Research Report

연구보고서 RE-152

SW안전 패러다임의 변화와 기술 대응 동향 연구

A Study on changes in SW safety paradigm and its technological trends

추형석/유재흥/조원영/안성원/노재원

2023.6.

SPRi 소프트웨어정책연구소

이 보고서는 2022년도 과학기술정보통신부 정보통신진흥기금을 지원받아 수행한 연구결과로 보고서 내용은 연구자의 견해이며, 과학기술정보통신부의 공식입장과 다를 수 있습니다.

목 차

제1장 서론 ····· 1

제1절 연구의 배경 및 필요성 ····· 1
1. 연구의 배경 ····· 1
2. 연구의 필요성 ····· 2
제2절 연구의 수행 방법 ····· 3
1. 전문가 집단의 정성적 연구 ····· 4
2. 언론기사 빅데이터 분석 ····· 5
3. 연구의 흐름 ····· 6
제3절 연구의 구성 ····· 7

제2장 SW 신기술 안전 사고 조사 ····· 8

제1절 개요 ····· 8
1. SW 신기술 안전 사고 조사 방법 ····· 8
2. SW 신기술 안전 사고 조사 양식 ····· 9
제2절 전문가 집단의 정성적 연구를 통한 SW 신기술 안전 사고 조사 ····· 11
1. SW 신기술 안전 사고 선정기준 ····· 11
2. SW 신기술 안전 사고 목록 ····· 12
제3절 언론기사 빅데이터 분석을 통한 SW 신기술 안전 사고 조사 ····· 21
1. 언론기사 빅데이터 분석 개요 ····· 21
2. 언론기사 빅데이터 수집 및 전처리 ····· 22
3. 언어모델을 활용한 언론기사 빅데이터 정제 ····· 28
4. 텍스트 마이닝을 활용한 언론기사 빅데이터 분석 ····· 32
5. SW 신기술 안전 사고 목록 ····· 43
제4절 SW 신기술 안전 사고 조사 결과 ····· 49

제3장 SW 신기술 위험원 유형화 모델 ····· 51

제1절 개요 ····· 51

1. SW 신기술 위험원 유형화 고려사항 ····· 51
2. SW 신기술 위험원 유형화 전략 ····· 52
제2절 선행연구 분석 ····· 54
1. SW 결함 분류 ····· 54
2. SW 실패를 고려한 연구 ····· 60
3. ICAM 사고 원인분석 방법론 ····· 67
제3절 SW 신기술 사고 유발 위험원 유형화 ····· 70
제4절 SW 신기술 안전 사고 조사결과의 적용 ····· 75
1. SW 신기술 사고 유발 위험원 유형화와 대응 ····· 75
2. SW 신기술 사고 유발 위험원 유형화와 대응 결과 ····· 77
제4장 SW 신기술 안전 대응 동향 ····· 79
제1절 개요 ····· 79
제2절 인공지능 ····· 79
제3절 메타버스 ····· 84
제4절 블록체인 ····· 88
제5장 결 론 ····· 98
[별첨] SW 신기술 사고 조사 보고서 목록 ····· 103
참고문헌 ····· 198

표 목 차

<표 2-1> 언론기사 수집을 위한 팩티바 언론사 목록 ····· 23
<표 2-2> 언론기사 데이터 수집 조건 ····· 24
<표 2-3> AI 분야 검색 키워드 ····· 24
<표 2-4> 메타버스/블록체인 분야 검색 키워드 ····· 25
<표 2-5> 데이터 프레임 ····· 27
<표 2-6> 예외처리를 통한 분석 데이터 현황 ····· 27
<표 2-7> 언어모델 재학습을 위한 학습 데이터 ····· 30
<표 2-8> 재학습된 언어모델 결과 ····· 31
<표 2-9> AI 분석 대상 데이터 토픽 모델링 결과 - 7대 토픽별 키워드 ····· 34
<표 2-10> 메타버스/블록체인 분야 토픽 모델링 결과 - 6대 토픽별 키워드 ····· 39
<표 2-11> SW 신기술 안전 사고 목록 (총 37건) ····· 49

<표 3-1> 기존 SW 결함 분류 스키마 개요 (Ploski et al., 2007) ····· 54
<표 3-2> SW-기능 실패 모드 분류 체계 ····· 62
<표 3-3> SW 기능의 타이밍 관련 실패 모드에 대한 분류 ····· 63
<표 3-4> SW 기능의 값 관련 실패 모드에 대한 분류 ····· 63
<표 3-5> SW 기능의 상호작용 실패 모드에 대한 분류 ····· 64
<표 3-6> SW 기능 수준에 관련된 실패 모드 유형 연구들 ····· 66
<표 3-7> 관련 연구들의 위험원 커버리지 비교 분석 ····· 74
<표 3-8> SW 신기술 사고 37건의 SW 신기술 사고 유발 위험원 유형 대응 결과 ····· 77

<표 4-1> SC42의 세부 조직 ····· 81
<표 4-2> DMA 5조 : 케이트키퍼가 준수해야 하는 의무 ····· 86

그 림 목 차

[그림 1-1] 연구 수행 방법의 개괄적 흐름 ····· 6

[그림 2-1] 안전사고 보고서 양식 ····· 10
[그림 2-2] 언론기사 빅데이터 분석의 흐름 ····· 22
[그림 2-3] 언론기사를 수집하는 331개 언론사의 분포 ····· 23
[그림 2-4] 팩티바에서 수집된 언론기사의 예시 ····· 26
[그림 2-5] AI 분야 분석 대상 데이터의 워드 클라우드 ····· 33
[그림 2-6] 키워드 네트워크 생성 규칙과 시각화 ····· 35
[그림 2-7] AI 토픽 1 - 자율주행 관련 사고 ····· 35
[그림 2-8] AI 토픽 2 - 안면인식 오남용과 감시 체계 ····· 36
[그림 2-9] AI 토픽 3 - 딥페이크 관련 사건 사고 ····· 36
[그림 2-10] AI 토픽 4 - 알고리즘과 챗봇의 편향 문제 ····· 37
[그림 2-11] AI 토픽 5 - AI 역기능에 대한 문제제기 ····· 37
[그림 2-12] AI 토픽 6 - AI 관련 규제의 필요성 ····· 38
[그림 2-13] AI 토픽 7 - AI 보안 이슈 ····· 38
[그림 2-14] 메타버스/블록체인 분야 분석 대상 데이터의 워드 클라우드 ····· 40
[그림 2-15] 블록체인 토픽 1 - 블록체인 관련 범죄 ····· 40
[그림 2-16] 블록체인 토픽 2 - DeFi 논란 ····· 41
[그림 2-17] 블록체인 토픽 3 - 블록체인 규제의 필요성 ····· 41
[그림 2-18] 블록체인 토픽 4 - 암호화폐의 위험성 ····· 42
[그림 2-19] 블록체인 토픽 5 - NFT 관련 이슈 ····· 42
[그림 2-20] 메타버스 토픽 - 메타버스 관련 이슈 ····· 43

[그림 3-1] SW-ICAM의 논리적 개발 흐름 ····· 52

[그림 3-2] SW 기능과 모든 수준의 기반 컴포넌트 ····· 62
[그림 3-3] Thieme 외 연구에서 식별한 SW 실패 모드 도식화 ····· 66
[그림 3-4] ICAM 절차 ····· 68
[그림 3-5] ICAM에서 제시한 14가지 organizational factor type ····· 69
[그림 3-6] SW 신기술 사고 유발 위험원 유형화 ····· 70
[그림 3-7] SW 신기술 사고 유발 위험원 대분류의 비중 ····· 78

[그림 4-1] AI ACT 설명 ····· 83

요 약 문

1. 제 목 : SW안전 패러다임의 변화와 기술 대응 동향 연구

2. 연구 목적 및 필요성

그간 SW 안전은 사람의 신체적, 물리적 피해를 유발할 수 있는 분야를 중심으로 발전해 왔다. 그러나 SW의 영향력은 급격하게 확장되고 있다. SW 신기술로 볼 수 있는 인공지능, 블록체인, 메타버스 등은 세상을 혁신시키는 핵심 기술로 부상하고 있으나, 그 이면에는 다양한 사건 사고가 존재한다. 이러한 사고는 비단 신체적, 물리적 피해를 넘어 정신적, 경제적 피해를 유발시키고 있다.

이제 SW 안전은 전통적인 SW 안전의 영역에서 탈피하여 디지털 전환시대의 총체적인 안전을 확보하기 위한 수단으로 변화하는 길목에 서 있다. 본 연구의 목적은 SW 신기술로 인해 발생한 주요 사고를 예방 및 완화하기 위해 SW 안전의 역할을 모색하는 것이다. 그 첫걸음은 SW 안전 프로세스의 첫 단추인 위험원 분석으로 출발한다. 특히 본 연구는 인공지능, 블록체인, 메타버스의 3대 SW 신기술 분야에서 잠재적인 위험원 유형화를 통해 SW 안전 패러다임의 변화를 도출하고자 한다.

3. 연구의 구성 및 범위

본 연구는 크게 △ SW 신기술 안전 사고 수집, △ SW 신기술 사고 유발 위험원 유형화, △ SW 신기술 안전 사고의 위험원 유형 대응으로 구성된다. 이후에는 SW 신기술과 관련된 SW 안전 동향에 대해서 살펴볼 것이다.

SW 신기술 안전 사고 수집은 크게 두 가지 접근을 취한다. 첫 번째는 SW 안전 전문가 집단의 정성적 연구이다. 정성적 연구에서는 SW 안전의 관점에서 유의미한 SW 신기술 안전 사고를 수집한다. 두 번째는 언론기사 빅데이터 분석을 통해 정성적 연구의 객관성을 보완한다. 언론기사 데이터는 2017년 1월 1일부터 66개월간 수집된 SW 신기술 관련 언론기사 약 35만 여건을 대상으로 한다. 이 언론기사 빅데이터에서 SW 신기술 관련 사고, 한계 및 문제점을 언급하는 단락을 언어 모델 재학습을 통해 도출하고, 텍스트 마이닝 기법을 활용하여 주요 사건을 도출한다.

SW 신기술 사고 유발 위험원 유형화는 다양한 선행 연구 분석과 융합을 수행한다. 전통적인 SW 안전 분야에서 다루고 있는 SW 결함 모델, SW 실패 모델과 더불어 사고조사 방법론, SW 신기술의 특성까지 반영하여 복합된 모델을 제안한다.

SW 신기술 안전 사고는 SW 신기술 사고 유발 위험원 유형에 대응시켜 어떠한 위험

원이 주로 나타나는 지를 파악한다. 이 때 사고의 내용과 위험원 유형의 대응은 SW 안전 전문가의 전문성을 바탕으로 추정했기 때문에 한계가 있을 수 있다.

마지막으로 SW 안전의 관점에서 SW 신기술에 대응하는 동향을 가용한 자료를 토대로 주요국의 정책 및 법제화 동향, 글로벌 표준 동향으로 정리하여 SW 안전의 현재를 가늠해보고자 한다.

4. 연구 내용 및 결과

SW 신기술 안전 사고는 총 37건이 수집되었다. 이중 인공지능이 18건, 메타버스가 9건, 블록체인이 10건을 차지했다. SW 안전 전문가 집단의 정성 연구에서는 26건이 도출되었고, 언론기사 빅데이터 분석에서는 25건이 도출되었는데, 중복되는 14건을 제외한 11건을 추가하여 총 37건의 SW 신기술 안전 사고를 확보했다. SW 신기술 안전 사고는 영국의 IT 인프라 라이브러리의 사고 조사 양식을 준용하여 보고서화 했다.

SW 신기술 사고 유발 위험원 유형은 7대 대분류의 38가지 항목으로 구성했다. 이를 도출하기 위해 영국의 사고 조사 방법론을 차용하여 SW 결함과 실패 모델을 융합했다. SW 신기술 분야에 대한 SW 안전의 위험원 유형은 선행연구가 부재하여 부득이하게 포괄적으로 정의했다.

SW 신기술 안전 사고는 SW 안전 전문가의 식견을 통해 SW 신기술 사고 유발 위험원 유형에 대응된다. 그 결과 SW 안전 메커니즘 부재, SW 안전 제도 미흡 등의 항목이 높게 관측되어, SW 안전의 인식 자체가 매우 부족한 것으로 나타났다. 특히 SW 신기술은 악용과 오남용으로 인한 위험원도 매우 높게 측정되어 시급한 대처가 필요할 것으로 판단된다.

SW 안전 관점에서의 SW 신기술에 대한 실효성 있는 대안은 전무한 상황이다. 인공지능 분야에서는 규제 법안 제정과 표준 활동이 있으나 실제로 적용되기까지는 아직 많은 시간이 필요할 것으로 판단된다.

5. 기대효과

본 연구는 SW 안전의 관점에서 SW 신기술에 적용시켜본 실험적인 연구라고 볼 수 있다. SW 안전에서의 위험원 분석은 사고 사례 하나의 위험원을 도출하기 위해서 소스 코드 분석부터 사고의 재현까지 매우 광범위한 테스트를 수행한다. SW 신기술은 단편적인 사고 내용만으로 이루어져 있어 위험원 분석이 사실상 불가능했으나, 위험원 유형화라는 관점으로 추상화시킨 것이 본 연구의 큰 기여 중 하나이다.

본 연구는 SW 안전이 디지털 안전으로의 패러다임 변화를 위해 기여할 수 있는 시작을 마련했다는 점에서 앞으로 지속적인 연구가 필요할 것이다.

SUMMARY

1. Title: A Study on changes in SW safety paradigm and its technological trends

2. Purpose and Necessity of the Research

SW safety has been developed focusing on areas which can cause physical and physical harm to people. However, the influence of SW is rapidly expanding. Artificial intelligence, blockchain, and metaverse, which can be seen as emerging SW, are arising as core technologies that innovate the world, but there are various incidents and accidents behind them. These incidents cause not only physical harm, but also psychological harm.

Now, SW safety is on the way to change into a means to secure overall safety in the era of digital conversion, breaking away from the realm of traditional SW safety. The purpose of this study is to explore the role of SW safety in preventing and mitigating major incidents caused by emerging SW. The first step starts with hazard analysis, which is the first step in the SW safety process. In particular, this study aims to derive a change in the SW safety paradigm by categorizing potential hazard in the three major emerging SW such as artificial intelligence, blockchain, and metaverse.

3. Composition and Range

This study is largely composed of △ collection of safety incidents in emerging SW, △ categorization of hazard that cause incidents in emerging SW, and △ a mapping to category of hazard in incidents in emerging SW. In the following, we will examine SW safety trends related to emerging SW.

A collection of incidents in emerging SW takes two approaches. The first is a qualitative study by a SW safety expert group. In the qualitative study, incidents in emerging SW that are meaningful from the viewpoint of SW safety are collected. The second is to supplement the objectivity of qualitative research through big data analysis of news articles. The news article data covers about 350,000 related to emerging SW collected for 66 months from January 1, 2017. In this big data,

paragraphs mentioning incidents, limitations and problems related to emerging SWare derived through pretrained language model, and major incidents are derived using text mining techniques.

The category of hazards that cause incidents by emerging SW is analyzed and converged with various preceding studies. In addition to the SW defect model and SW failure model, which are dealt with in the traditional SW safety field, we propose a complex model that reflects the incident investigation methodology and the characteristics of emerging SW.

Incidents in emerging SW are mapped to the category of hazard. At this point, since the mapping has been done by experts of SW safety, there will be some limitations on the results.

Finally, from the viewpoint of SW safety, based on the available data, the trend of emmerging SW is summarized into policies and legislation trends of major countries and global standard trends to gauge the current state of SW safety.

4. Main Contents and Results

A total of 37 cases of emerging SW were collected. Among them, artificial intelligence accounted for 18 cases, metaverse for 9 cases, and block chain for 10 cases. In the qualitative study of the SW safety expert group, 26 cases were derived, and in the big data analysis, 25 cases were derived with 14 overlappings. The incidents in emerging SW were reported by applying the incident form of ITIL.

The category of hazard causing incidents by emerging SW were composed of 38 items in 7 major categories. To derive this, the British incident investigation methodology was adapted and the SW defect and failure model was fused. The category of hazards of SW safety in emerging SW was inevitably broad due to the absence of prior studies.

The mapping between the list of incidents by emerging SW and the category of hazard has been purseued by SW safety experts. As a result, items such as absence of SW safety mechanism and insufficient SW safety system were highly observed, indicating that awareness of SW safety itself was very insufficient. In particular, emerging SW is judged to require urgent action as the hazard caused by abuse and misuse are measured very high.

In terms of SW safety, there is no effective ways to prevent or mitigate incidents

caused from emerging SW. In the field of AI, there are regulatory legislation and standard activities, but it will still take a long time to be actually applied.

5. Expected Effects

This study can be seen as an experimental study applied to emerging SW from the viewpoint of SW safety. Hazard analysis in SW safety performs a very wide range of tests from source code analysis to reproduction of an accident in order to derive a hazard from an incident. Hazard analysis was actually impossible since emerging SW consisted of very little information, but one of the major contributions of this study was to abstract it from the perspective of categorization.

In this study, continuous research will be needed in the future in that SW safety has prepared a start that can contribute to the paradigm change to digital safety.

CONTENTS

Chapter 1. Introduction ····· 1

Section 1. Backgrounds and Needs for Research ····· 1
1. Backgrounds for Research ····· 1
2. Needs for Research ····· 2

Section 2. Method of Research ····· 3
1. Qualitative Study by a Group of Experts ····· 4
2. Analysis of Article Big-data ····· 5
3. Flow of Research Methodology ····· 6

Section 3. Structure of Research ····· 7

Chapter 2. Investigation of Incidents in Emerging SW Technology ····· 8

Section 1. Outline ····· 8
1. Method of Investigation ····· 8
2. Form of Investigation ····· 9

Section 2. Investigation of Incidents in Qualitative Study ····· 11
1. Ranges and Rules ····· 11
2. List of Incidents ····· 12

Section 3. Investigation of Incidents in Big-data Analysis ····· 21
1. Outline of Article Big-data Analysis ····· 21
2. Collection and Preprocessing of Article Big-data ····· 22
3. Article Big-data Cleansing by Language Model ····· 28
4. Article Big-data Analysis by Text Mining ····· 32
5. List of Incidents ····· 43

Section 4. Result of Incidents ····· 49

Chapter 3. A Categorization of Hazard in Emerging SW Technology ····· 51

Section 1. Outline ····· 51
1. Considerations for Categorization ····· 51
2. Strategy for Categorization ····· 52

Section 2. Related Studies ····· 54
1. SW Error Model ····· 54
2. SW Failure Model ····· 60
3. ICAM Model ····· 67

Section 3. A Categorization of Hazard in Emerging SW Technology ····· 70

Section 4. Application to the Categorization of Incidents ····· 75
1. A Mapping between Categorization and Incidents ····· 75
2. Results of Mapping ····· 77

Chapter 4. Trends in Safety of Emerging SW ····· 79

Section 1. Outline ····· 79

Section 2. Artificial Intelligence ····· 79

Section 3. Metaverse ····· 84

Section 4. Blockchain ····· 88

Chapter 5. Conclusion ····· 98

[Attatchment] List of Incidents in Emerging SW ····· 103

Reference ····· 198

제1장 서 론

제1절 연구의 배경 및 필요성

1. 연구의 배경

소프트웨어(Software, 이하 SW)는 디지털 사회를 지탱하는 근간이다. 특히 SW는 기술적 혁신과 사회 변화에 민감하게 반응하여 다채롭게 변화하며 진화하고 있다. 예를 들면 지능의 자동화를 구현한 SW인 인공지능(Artificial Intelligence, 이하 AI)은 빅데이터를 학습하는 딥러닝의 출현으로 사람 수준의 언어를 구사하고, 사물을 인식할 수 있으며, 다양한 콘텐츠를 생성할 수도 있다. 또한 탈중앙화를 활용해 데이터 보안의 혁신을 일으킨 블록체인은 디지털 자산의 거래라는 새로운 지평을 열었다. 코로나19로 인한 비대면 사회가 지속되자 가상 현실(Virtual Reality)은 메타버스로 진화하여 가상 세계의 확산이 본격화되었다.

그러나 혁신적인 SW 신기술이 속속 등장하여 세상을 편리하게 바꾸는 만큼, 이로 인한 부작용도 적지 않다. 특히 SW 신기술로 인한 사고는 법적 테두리 밖에 있는 경우가 많아 사고의 재발방지가 매우 어렵다. 이에 대응하기 위해 국제사회는 표준을 제정하고 원칙이나 가이드라인을 공개하고 있으나 SW 신기술의 발전 속도에 발맞춰 가기에는 어려움이 많다. 가장 대표적인 SW 신기술인 AI의 경우에도 유럽에서 규제법안이 나온 상황이지만 실효성에 대한 비판이 제기되고 있는 실정이다.

그렇다면 새롭게 등장하는 SW 신기술의 사건 사고를 예방하기 위한 대안이 필요할 것이다. 특히 SW 신기술이 법제화의 영역에 들어서기 전까지 SW 신기술과 관련된 모든 주체들이 다각적인 노력을 펼쳐야 한다. 본 연구에서는 그 노력의 시작을 SW 공학의 한 분야인 SW 안전(SW Safety)에서 찾아보고자 한다.

그간 SW 안전은 자동차, 항공, 철도 등 SW 오류나 실패로 인해 막대한 피해가 예상되는 주요 분야를 중심으로 발전해 왔다. 그 이유는 SW가 도메인의 특성에 맞춰 유연하게 변화하기 때문이다. 이렇게 SW 안전은 각 분야별로 특화된 표준과 민간 중심의 자발적 참여로 발전해 왔는데, 최근 디지털 전환의 핵심 기술로 SW의 위상이 점차 높아짐에 따라 범용적인 관점에서의 SW 안전에 대한 관심이 더욱 높아지고 있는 상황이다.

2. 연구의 필요성

그러나 SW 안전은 SW 신기술의 변화에 민첩하게 대응하기에는 어려운 요인들이 산재하고 있다. SW 안전 프로세스는 SW 안전 요구사항을 도출하고 이를 구현하는 것을 중심으로 이루어지나, SW 안전 요구사항을 도출하기 위해서는 SW관련 사고가 발생한 위험원 분석과 위험 평가가 뒷받침되어야 하기 때문이다. 여기서 SW 안전 프로세스의 시작이라고 할 수 있는 위험원 분석(Hazard Analysis)은 소스코드 레벨의 SW 사고 분석 보고서가 뒷받침되어야 한다는 점에서, 현재 논의되고 있는 SW 안전이 SW 신기술에 직접 적용되기에는 요원하다고 볼 수 있다.

우리나라에서는 지난 2020년 12월 SW진흥법이 전면 개정됨과 동시에 SW 안전이 법적으로 정의되었다. SW진흥법 제2조 8항에서는 SW안전을 “외부로부터의 침해행위가 없는 상태에서 SW 내부적인 오작동 및 안전기능(사전 위험분석 등을 통하여 위험 발생을 방지하는 기능을 말한다) 미비 등으로 인하여 발생할 수 있는 사고로부터 사람의 생명이나 신체에 대한 위험에 충분한 대비가 되어 있는 상태를 말한다.”로 정의하고 있다. 이 정의에서는 사람의 생명이나 신체에 대한 피해를 명시하고 있다는 점에서 한편으로는 협소한 정의라고도 볼 수 있는데, 현재 발발하고 있는 SW 신기술과 관련된 사건 사고는 비단 물리적·신체적 피해 뿐만 아니라, 경제적·정신적 피해까지 일으키고 있기 때문이다.

이러한 상황에서 SW 안전은 협소한 정의에서 벗어나 더 범용적인 관점에서 SW를 포괄할 수 있는 패러다임의 변화가 필요하다. 즉, 사후적인 관점에서 SW 안전 요구사항을 도출하는 것에서 벗어나 SW 신기술로 인한 사건 사고를 예방하는 차원에서 SW 안전의 역할을 찾아야 한다.

이에 본 연구에서는 SW 안전의 패러다임 변화와 이에 대응할 수 있는 SW 안전의 방향을 도출하고자 한다. 특히 다양한 SW 신기술 중에서 최근 각광을 받고있는 AI, 메타버스, 블록체인의 3대 분야를 선정하여 SW 신기술로 인한 주요 사건 사고를 수집했다. 그 방법으로는 전문가 집단에 의한 정성적 연구와 언론기사 빅데이터 분석을 활용했다. 또한 특정 사건 사고에 대한 소스코드 레벨의 조사 보고서가 부재한 상황을 고려하여 위험원 분석의 대안으로 위험원 유형화(Hazard Categorization)을 도출했다. 이는 가칭 SW 사고 유발 위험원 유형으로 명명하여 각 사건 사고에 대한 위험원 유형을 대응시켜 SW 안전의 패러다임 변화에 대해 논의하고자 한다.

제2절 연구의 수행 방법

본 연구에서는 SW 신기술로 인한 SW 안전의 패러다임 변화를 도출하기 위한 가장 첫걸음으로 SW 신기술로 인해 발생한 사건 사고의 위험원을 유형화하는 접근을 취한다. 이를 위해서 연구의 수행 과정을 크게 4단계로 나누어 볼 수 있다.

① SW 신기술 안전 사고 수집

SW 신기술 안전 사고는 주로 뉴스 기사를 통해 전달되는 경우가 많으며, 자율주행차와 같이 사람의 신체적 피해가 자명한 경우에는 정부 기관의 조사보고서가 발간되고 있다. 본 연구에서는 SW 신기술 사고를 수집하기 위해 두 가지 수행 방법을 고려했다. 첫 번째는 SW 안전 전문가 집단이 SW 안전 사고의 관점에서 유의미한 것을 정성적인 관점에서 수집하는 것이다. 두 번째는 언론기사 빅데이터 분석을 통해 가장 주요하게 언급된 SW 신기술 안전 사고를 수집하는 방법을 택했다. 이 두 가지 방법으로 도출된 SW 신기술 안전 사고의 합집합을 최종 분석 데이터로 선정했다.

② SW 신기술 사고 유발 위험원 유형화

SW 신기술 안전 사고 데이터를 활용하여 SW 안전의 패러다임 변화상을 도출하기 위해서는 이러한 사고들이 발생한 위험원을 유형화해야 한다. 본 연구에서는 전문가 집단의 문헌 연구를 통해 SW 신기술 안전 사고 데이터를 포괄할 수 있는 가칭 SW 신기술 사고 유발 위험원 유형을 도출했다. 특히 SW 안전에서 사고의 원인으로 바라볼 수 있는 SW 오류와 SW 실패 모델을 비교 분석하고, SW 신기술의 기술적 특성을 고려했다.

③ SW 신기술 안전 사고의 SW 신기술 사고 유발 위험원 유형에의 대응

그 다음 단계는 SW 신기술 안전 사고가 어떠한 SW 신기술 사고 유발 위험원 유형에 속하는지를 대응시키는 것이다. 이 대응 결과를 토대로 각 위험원 유형에 대한 빈도를 도출할 수 있다. 이 빈도는 SW 신기술의 다양한 사건 사고에 대응하기 위한 SW 안전의 우선순위를 도출할 수 있다는 점에서 의의를 찾을 수 있으며 SW 안전의 패러다임 변화와 그 대응을 논의할 수 있는 기반자료로 활용된다.

④ SW 신기술의 안전 확보를 위한 대응 동향 조사

마지막은 SW 신기술별 안전 확보를 위한 주요국의 정책 동향과 더불어 표준화 및 법제화 현황을 분석하여 기술 대응 동향을 종합적으로 분석한다.

1. 전문가 집단의 정성적 연구

SW 신기술에 대한 SW 안전은 그간 논의에 대한 필요성은 인지하고 있었으나, SW 신기술 개발에 SW 안전 프로세스를 직접 적용하기에는 매우 어려운 상황이다. 그 이유는 앞서 논의한대로 SW 신기술로 인한 안전 사고에 대한 원인 규명을 담은 분석 보고서가 부재하고, 이에 따라 SW 안전 요구사항 자체를 도출하기가 사실상 불가능하다는 점이다.

그럼에도 최근 사건 사고 보고서에서는 SW 안전 메커니즘 혹은 안전 문화의 정착을 강조하고 있다. 2018년 3월 미국 애리조나주 탬피에서 자율주행모드로 시험 운행중인 우버의 자율주행차가 최초로 보행자를 사망시키는 사건이 발생하여 세간에 큰 충격을 남겼다. 이 사건을 지속적으로 조사한 미국 연방교통안전위원회에서는 자율주행차를 보조하는 운전자의 의무에 대한 책임을 묻는 한편, 자율주행차의 시험 운행에 대한 안전 문화 미숙이라는 점을 주요 원인으로 지목했다(NTSB, 2019).

이렇게 SW 안전의 현실과 확대 필요성의 간극을 메우기 위해서는 SW 안전을 잘 이해하고 있는 전문가 집단의 연구가 필요하다. 본 연구에서는 우리나라를 대표하는 SW 안전 전문가로 구성된 소프트웨어와사회안전협회와 협력하여 연구를 수행했다.

전문가 집단의 정성적 연구(이하 정성 연구)에서는 가장 핵심적인 과업으로 SW 신기술 사고 유발 위험원 유형화 연구를 수행했다. 위험원 유형화는 본 연구의 전개에서 중심을 차지하고 있는 것으로, 수집한 SW 신기술 안전 사고 보고서가 어떠한 위험원 유형으로 구분되는가를 확인할 수 있는 틀(frame)이다. 특히 SW 신기술은 SW 안전에서 바라보는 SW와는 상이한 점이 많기 때문에, 레거시를 포괄하는 위험원 유형화가 필요하다.

또한 전문가 집단의 중요한 역할 중 하나는 언론기사 빅데이터 분석과 정성적 연구에서 도출된 SW 신기술 안전 사고를 SW 신기술 사고 유발 위험원 유형에 대응시키는 것이다. 이것은 전문가의 전문 지식을 바탕으로 추정하는 정성적 접근이외에는 적합한 수단이 없기 때문에, 연구 결과의 한계와도 맞물려 있다. 그러나 SW 신기술과 SW 안전의 관계를 기술하는 선행연구가 매우 부족한 상황이기 때문에, 객관적 지표나 정량적 방법론을 적용하기 어려운 점을 고려해야 한다.

마지막으로 정성적 연구에서는 SW 안전의 관점에서 SW 신기술에 대한 기술 대응 동향을 주요국의 정책적 대응과 표준·법제화 동향으로 정리했다.

2. 언론기사 빅데이터 분석

SW 신기술과 관련된 사고가 전파되는 가장 빠른 루트는 언론기사다. 따라서 언론기사를 수집하여 주요 SW 신기술 안전 사고를 수집한다는 접근은 정성 연구에서 경험적으로 수집한 SW 신기술 안전 사고를 보완하기에 매우 적절한 접근방법일 것이다.

본 연구에서는 언론기사 데이터를 수집하기 위해 영국 다우 존스에서 운영하는 플랫폼인 팩티바(Factiva)를 활용했다. 팩티바는 전 세계 약 3.3만 여개의 언론사에서 생산하는 언론기사를 운영하는 대형 플랫폼으로, 본 연구에서는 2017년 1월 1일부터 66개월 간 주요 글로벌 언론사 331개에서 생산한 AI, 메타버스, 블록체인 관련 영문기사 35만 건을 수집했다.

언론기사 빅데이터 분석의 목적은 35만 건의 언론기사 데이터에서 어떠한 기사가 SW 신기술의 사고를 언급하고 있는지를 찾아내는 것이다. 그러나 SW 신기술 사고에는 사고 자체의 경중과 피해도가 상이하기 때문에, 단발적이고 경미한 피해가 발생하는 경우는 배제할 필요가 있다. 이에 본 연구는 텍스트 마이닝을 통해 SW 신기술 사고를 유형화하고, 유형별로 가장 많이 언급된 사고를 추출하여 SW 신기술 안전 사고 보고서를 도출했다.

SW 신기술 사고 유형화를 위해서는 분석 대상 데이터를 단락으로 선정했다. 그 이유는 언론기사의 특성에서 찾을 수 있는데, 언론기사는 기본적으로 사실전달의 중립적인 내용이 대부분이기 때문이다. 본 연구에서는 35만 건의 언론기사에서 도출된 670만 개의 단락을 대상으로, 특정 단락이 SW 신기술과 관련된 사건 사고, 한계, 문제점을 언급하는지 여부를 판단하는 것을 첫 번째 과업으로 삼았다.

이 첫 번째 과업을 수행하기 위해, 본 연구에서는 언어모델 재학습 방법을 활용했다. 언어모델(Language Model)은 자연어 처리의 한 방법론으로 주어진 단어나 문장 뒤에 올 단어를 예측하는 일종의 생성 모델이다. 언어모델은 최근 AI 발전으로 괄목한 성장을 거두어 사전학습 언어모델(Pre-trained Language Model)이라는 트렌드를 만들고 있다. 이것은 고성능 언어모델을 특정과업을 수행할 수 있게 재학습[1]을 시킨다는 접근으로 해석할 수 있다. 언어모델 재학습을 위해서는 재학습 데이터가 필요하다. 이를 위해 본 연구에서는 특정 단락이 SW 신기술의 사건 사고, 한계, 문제점과 관련이 있는지 여부를 라벨링하여 재학습 시키는 과정을 거쳤다.

1) 언어모델을 확장하여 생각해보면, “A와 B는 비슷합니까? ___ ”라는 문장에서 빈칸에 들어갈 단어의 예측은 A와 B가 비슷한지 여부를 판단할 수 있어야 한다. 따라서 상식이 풍부한 언어모델의 경우 단순한 질의 응답부터 기계 번역 등 다양한 자연어 처리 과업에 적용시킬 수 있다.

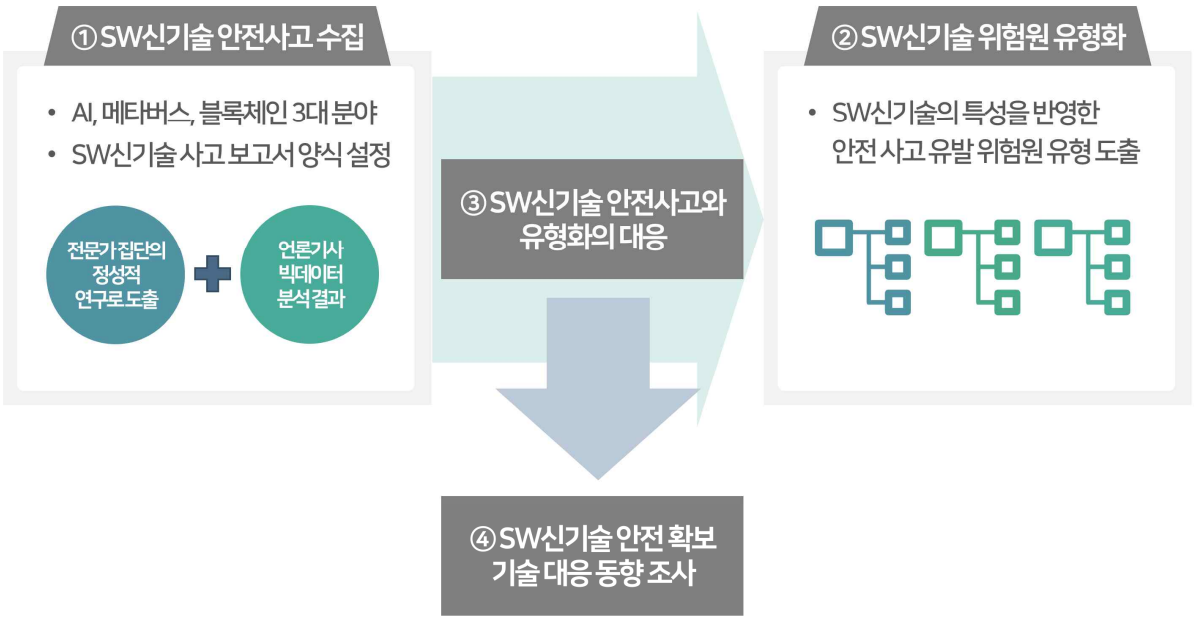
이후 재학습이 완료된 언어모델에 수집된 언론기사 단락을 대입시켜 SW 신기술의 사건 사고, 한계, 문제점 단락(이하 SW 신기술 사고 관련 단락)을 추출하게 된다. 이렇게 정제된 SW 신기술 사고 관련 단락은 텍스트 마이닝의 입력 데이터로 활용된다.

텍스트 마이닝은 정제가 완료된 SW 신기술 사고 관련 단락을 대상으로 한다. 먼저 SW 신기술 사고의 유형화를 위해서 LDA(Latent Dirichlet Allocation) 기반 토픽 모델링을 수행했다. 토픽의 수는 토픽별로 할당된 상위 키워드의 적절성을 보고 판단한다. 이후 각 토픽에 해당하는 단락을 별도로 추출하여 키워드 네트워크 분석을 수행한다. 이는 토픽에서 가장 중요한 사건을 도출하는 것으로 SW 신기술 사고 보고서의 주제를 선정하는데 활용했다.

언론기사 빅데이터 분석으로 도출한 결과는 두 가지이다. 첫 번째는 SW 신기술 안전 사고 목록을 확보함에 있으며, 두 번째는 빅데이터 분석 결과 자체로써의 시사점이다. 특히 본 연구에서의 언론기사 빅데이터 분석은 영문에 한정되었지만 가용한 언론기사를 모두 수집해봤다는 점에서 의미가 있다.

3. 연구의 흐름

총 4단계로 이루어진 연구의 전반적인 절차는 다음 [그림 1-1]과 같다.



[그림 1-1] 연구 수행 방법의 개괄적 흐름

제3절 연구의 구성

본 연구는 다음과 같은 구성으로 전개된다.

먼저 2장에서는 SW 신기술 안전 사고 조사에 대한 내용을 다룬다. 이는 [그림 1-1]에서 ① SW 신기술 안전 사고 수집에 해당하는 내용이다. 먼저 1절에서는 SW 신기술 안전 사고 조사를 기술하는 양식의 선정과 세부 내용에 대해서 논의할 것이다. 이후 2절에서는 정성적 연구에서는 SW 안전 전문가들이 선정한 SW 신기술 사건 사고를 수집한 결과를 다룬다. 이러한 접근은 개인의 주관이나 성향에 민감하게 반응할 수 있기 때문에, 이를 보완하기 위한 방법으로 3절에서 언론기사 빅데이터 분석 내용을 논의한다. 정성적 연구와 빅데이터 분석으로 도출한 SW 신기술 안전 사고 조사의 결과와 시사점에 대한 내용을 4절에서 논의하는 것으로 2장을 마무리 한다.

3장에서는 SW 신기술 위험원 유형화 모델을 다루며, 이 내용은 [그림 1-1]의 ② SW 신기술 위험원 유형화에 해당하는 내용이다. 1절에서는 SW 신기술 위험원 유형화를 어떻게 정의하고 논의를 전개할 지에 대한 고려사항을 다룬다. 2절에서는 SW 신기술 위험원 유형화를 위한 선행연구 분석으로 △ SW 결함 분류, △ SW 실패를 고려한 연구, △ 사고 원인분석 방법론(ICAM)의 세 가지 항목으로 나누어 볼 것이다. 3절에서는 선행연구를 바탕으로 정의한 SW 신기술 사고 유발 위험원 유형화에 대해서 논의할 것이다. 마지막 4절에서는 [그림 1-1]의 ③ SW 신기술 안전 사고와 유형화 대응을 다루며, 이를 통해 SW 신기술 위험원이 주로 발생하는 유형을 파악하여 마지막 결론에서 그 대응방향을 논의하겠다.

4장에서는 SW 신기술 안전에 대한 대응 동향을 다룬다. 이 장에서는 SW 안전의 관점에서 SW 신기술에 대응하는 동향이 있는지를 조사한 내용을 다룬다. 조사 범위는 AI, 메타버스, 블록체인 기술에 대응하는 SW 안전 관련 정책, 표준, 법률을 중심으로 설정했다. 먼저 1절에서는 SW 신기술 안전 대응 동향의 전반적인 조사 방법에 대해 논의하고 이어지는 2절, 3절, 4절에서는 각각 AI, 메타버스, 블록체인에 대한 주요국의 정책, 표준 및 법제화 동향, 시사점에 대한 내용을 다룰 것이다.

마지막으로 5장에서는 SW 신기술의 사건 사고에 대응할 수 있는 SW 안전의 패러다임 변화을 결론으로 논의하겠다. 또한 언론기사 빅데이터 분석 결과의 함의도 논의하여 SW 신기술 사건 사고, 한계 및 문제점 전반에 대한 내용을 종합적으로 살펴보겠다.

제2장 SW 신기술 안전 사고 조사

제1절 개 요

1. SW 신기술 안전 사고 조사 방법

SW 신기술 안전 사고 조사는 연구의 흐름을 표현한 [그림 1-1]에도 나와있는 바와 같이 전문가 집단의 정성 연구와 언론기사 빅데이터 분석 결과를 취합하여 도출하는 형태이다. SW 신기술 분야는 최근 각광받고 있는 기술의 성숙도 등을 고려하여 AI, 메타버스, 블록체인의 3대 분야로 제한했다. 특히 정성 연구에서는 SW 안전 전문가의 관점에서 SW 신기술 안전 사고를 도출하기 때문에 SW 신기술 전문가가 사고를 도출하는 것보다 공정성을 확보할 수 있다. 여기에 SW 안전 전문가의 전문성을 더하기 위하여 언론기사 빅데이터 분석 결과로 SW 신기술 안전 사고를 추가적으로 도출했다.

정성 연구에서는 기본적으로 국내외 온라인 및 오프라인 언론기사에서 발표된 자료들을 중심으로 조사했다. 단순한 사상 사고 이외에도 법적 분쟁, 재정적 손실 등 사고의 범위를 보다 광범위한 시각에서 조사하도록 했다. 언론기사 이외에 연구보고서, 조사 보고서 등을 참고했으며, 다양한 도메인에서 접근하는 것을 고려했다. 웹 검색에서 단순히 ‘사고’라는 단어로 사용하는데는 한계가 있으므로, 도메인별 사고 관련 단어(예 - Crash)들을 활용하여 검색했다. 또한 국외의 경우 incidentdatabse.ai(Artificial Intelligence Incident Database), statista.com과 같은 접근 가능한 Incident Database들을 활용했다. 그러나 사고 통계 데이터가 부재한 상황이므로 전문가 시각에서의 기술적 중요성, 언론기사의 빈도 등을 종합적으로 고려하여 SW 신기술 안전 사고를 선정했다.

언론기사 빅데이터 분석에서는 SW 신기술과 관련된 사건 사고, 한계 및 문제점을 언급하는 단락(Paragraph)의 텍스트 마이닝을 통해 SW 신기술 안전 사고를 도출하는 접근을 취했다. 이것은 크게 두 가지 단계로 이루어지는데, 첫 번째는 분야를 적시하기 위한 군집 알고리즘의 하나인 토픽 모델링을 적용하였다. 토픽 모델링은 언론기사 단락을 소수의 토픽으로 군집하는 것으로, 해당 토픽을 대표하는 키워드를 산출하게 된다. 이후 토픽 별로 할당된 단락에 대해 키워드 네트워크를 생성하고 분석하여 주요 사건을 도출하는데 활용되었다. 예를 들어 ‘자율주행차’ 토픽의 키워드 네트워크는

‘애리조나’, ‘탬피’, ‘충돌사고’, ‘사망’, ‘우버’ 등의 키워드가 주로 군집되어 2018년 3월 미국 애리조나 탬피에서 발생한 우버 자율주행차가 보행자를 사망시킨 사건으로 적시할 수 있다. 키워드 네트워크 분석에는 군집 알고리즘의 하나인 커뮤니티 분석을 활용했으며, 키워드를 통한 사건의 유추는 연구진의 논의를 통해 선정했다.

이렇게 두 가지 연구 방법을 통해 도출된 SW 신기술 안전 사고는 SW 신기술 안전 사고 조사 양식에 의거하여 보고서 형태로 정리되고, 각 사고의 원인이 되는 위험원이 SW 신기술 위험원 유형과 대응되어 어떠한 위험원이 가장 많이 존재하는 데 활용된다.

2. SW 신기술 안전 사고 조사 양식

사고 조사 양식은 사고에 대한 설명과 내용, 특성들을 나타낼 수 있도록 작성하여야 한다. 이를 위해 SW 공학 분야에서 사용되는 SW 장애 보고서의 활용을 고려했다. SW 장애 보고서 또한 다양한 양식들이 존재하는데 본 연구에서는 영국의 ITIL(IT Infrasturcture Library) 표준에 기반한 사고 보고서(Incident Report)를 변형하여 활용했다. 사고 조사 양식에는 다음과 같은 항목으로 구성한다.

- 사고 ID
- 사고 분야
- 사고 명
- 사고 보도 정보(사고자료 출처, 보도 날짜, 작성자)
- 사고 세부 정보(사고 국가, 사고 유형, 사고 일시, 사고 지속시간)
- 사고 내용
- 사고 피해 내역
- 잠재 원인 유형 추정에 대한 전문가 의견

SW 신기술 안전 사고 보고서 양식의 항목 중 ‘사고 지속시간’은 실제 사고가 지속된 시간을 의미한다. SW 서비스의 사고는 물리적 충돌과 같은 순간적인 사고 이외에 사고가 지속된 기간이 존재하기 때문이다. 사고 내용과 사고 피해 내역은 가능한 충분한 정보를 수집하여 최대한 충실히 기술하는 것을 전제로 하며, 불확실하거나 추정된 내용은 기술하지 않는다.

잠재 원인 유형 추정에 대한 전문가 의견은 SW 신기술 위험원 유형화 모델의 항목에 따라 관련성이 있는지 여부를 판단하고 이유를 기술한다. 또한 SW 신기술 사고 조사 보고서에 대한 전체적인 총평도 추가되었다. 잠재 원인 유형 추정은 정성 연구에 참여하는 전문가가 직접 수행하며, 전문가의 지식과 식견을 통해 가능한한 객관적으로 작성했다.

[그림 2-1]은 SW 신기술 안전 사고 조사 양식과 작성해야 하는 내용에 대한 개요를 나타낸다.

SW신기술 안전사고 보고서

No. SW신기술 분야-No.

사고 분야	AI / Metaverse / Blockchain 중 택1			
사고명	사고를 설명할 수 있는 사고명			
사고 보도 정보	사고자료 출처	사고 보고서의 출처		
	보도 날짜	사고 보도 날짜		
	작성자	보도자 성명		
사고 세부 내용	사고 국가(장소)	국가(지역)	사고 유형	사고의 종류
	사고 일시	사고발생년월일	사고 지속시간	문제지속시간
사고 내용	사고에 대한 요약 설명으로서 사고 시스템의 이름, 사고 시스템의 특성, 사고 경위, 관련 소송, 진행 결과 등에 대하여 기술한다.			
사고 피해 내역	보도된 사고 피해 대상자와 피해 내역, 관련 조치 등을 기술한다.			
잠재 원인 유형 추정에 대한 전문가 의견	해당 사고의 잠재 원인에 대하여 전문가 의견을 기술한다. 전문가 의견은 4장의 잠재 원인 유형에서 기술한 7가지 분야(Safety Mechanism, Actor, Environment, Organizational/Process/Policy, SW System, Data, Hardware) 관점에서 기술하도록 한다. 그리도 해당 사고 유형의 재발 방지를 위한 전문가 권고를 기술한다.			

[그림 2-1] 안전사고 보고서 양식

제2절 전문가 집단의 정성적 연구를 통한 SW 신기술 안전 사고 조사

1. SW 신기술 안전 사고 선정 기준

SW 신기술의 범위는 서론에서 논의했다시피 사건 사고 발발 이력과 기술의 성숙도 등을 종합적으로 판단하여 AI, 메타버스, 블록체인의 3대 분야로 한정했다. 정성 연구에서는 SW 안전 전문가가 언론 기사와 각종 보고서 등 문헌 분석을 통해 주요한 사건을 추출하는 과정을 거쳤다. SW 신기술 각 분야의 전문가를 통해 사건을 선별하는 것도 고려했지만, 부정적인 영역을 다룬다는 점에 있어서 객관성을 보장하기 어렵다. 즉, 정성 연구에서는 전문성보다는 객관성에 더 비중을 두고 사고를 선별했다고 볼 수 있다.

사고 조사 건수는 SW 신기술 분야의 특성을 고려하여 AI 10건, 메타버스 5건, 블록체인 5건으로 선별했다. 기간은 최근 3년간으로 한정하되 매우 중요한 사건인 경우 최근 5년간으로 정했다.

SW 신기술 안전 사고에서 “안전 사고”의 사전적인 기준은 SW 진흥법에 정의된 바와 같이 인명이나 신체에 사상을 입히는 사고로 한정되어 있다. 그러나 SW 신기술의 피해 양상을 고려해보면, “안전 사고”의 범위를 더 확장시켜 비단 신체적 상해뿐만 아니라 정신적 피해를 입힐 수 있는 직간접적인 사고를 포함시키고자 한다. 따라서 물리적 상해 이외에 경제적 손실부터 사회적 불안까지 야기할 수 있는 모든 사고를 안전 사고의 하나로 광범위하게 정의하고 진행했다.

정성 연구에서는 먼저 SW 안전 전문가의 관점에서 중요하다고 생각되는 사건들을 일차적으로 수집하고, 이후 총 20건의 SW 신기술 안전 사고를 선별했다. 선별 기준은 언론에 가장 많이 노출된 사고를 우선적으로 고려했다. 또한 사고 유형의 다양성을 확보하기 위해 가능한 유사한 사고는 배제했다. 정성 연구의 SW 신기술 안전 사고는 목표치인 20건 보다 많은 26건이 도출되었다. 각 분야별로는 AI가 14건, 메타버스가 7건, 블록체인이 5건으로 선택됐다.

2. SW 신기술 안전 사고 목록

정성 연구를 통해 도출된 26건의 SW 신기술 안전 사고의 주요 내용을 요약한 결과는 다음과 같다. 순서는 SW 신기술 분야(AI, 메타버스, 블록체인)-연번의 형태로 나열한다.

- AI-01
  - 사고명 : 신약 개발 AI의 악용 가능성 제기
  - 사고 내용 요약 : 미국의 신약개발업체 연구진은 실험을 통해 신약 개발 AI의 악용 가능성을 제기하였다. 메가신 이라는 머신러닝 알고리즘은 일반적으로 물질의 독성이 예측될 경우 감점하고, 반대로 생체적합성이 예측되면 점수를 더하는 식으로 적절한 약물 후보를 탐색하도록 학습된 AI인데, 이 방식을 살짝 바꿔 의도적으로 인체에 치명적인 독성물질을 찾고자 한 실험 결과 4만종의 독성물질을 찾아냈다. 연구진은 누구나 접근할 수 있는 공개 데이터베이스에 있는 정보로 AI를 훈련시켰으며, 이 실험결과에 따라 AI의 오용에 대한 진입 장벽이 낮은 현실에 우려를 표명하였다.

- AI-02
  - 사고명 : AI 딥페이크 기술을 이용한 여성 나체사진 유포
  - 사고 내용 요약 : AI 딥페이크 기술을 이용해 여성 이미지에서 옷을 지우고 나체로 합성하는 텔레그램 채널이 운영되고 있는 것으로 드러났다. 이 채널은 AI 기반 딥페이크 봇을 이용해 익명의 사용자가 옷을 입은 여성의 사진을 보내면 누드 버전의 사진을 받아볼 수 있게 하였다. 2019년 7월부터 1년간 가짜 나체 사진 유포로 피해를 입은 여성의 수는 약 10만 4852명으로 파악됐고, 조사 결과 일부 피해자는 미성년자로 확인되었다.

- AI-03
  - 사고명 : 자율주행차량에 의한 첫 보행자 사망사고
  - 사고 내용 요약 : 미국 애리조나서 직원이 탑승한 채로 자율주행 시험 중이던 우버 차량이 사람을 치어 숨지게 하는 사고가 발생했다. 횡단보도가 아닌 곳에서 자전거를 끌고 도로를 건너던 여성을 치어 숨지게 했으며, 이는 자율주행차

최초의 보행자 사망사고로 기록됐다. 차량 운전자는 직접 운전을 하지 않았기 때문에 책임이 없다고 주장했지만, 충돌 당시 스마트폰으로 스트리밍 방송을 보고있는 등 백업 운전자로서의 역할에 책임을 지지 않은 이유로 구속됐다.

- AI-04
  - 사고명 : 숨진 가족 목소리 되살리는 아마존 AI '알렉사'의 신기능 논란
  - 사고 내용 요약 : 1분 미만의 음성 샘플로 죽은 사람의 목소리까지 되살리는 알렉사의 신기능이 공개되면서 논란이 되었다. AI 목소리 모방 기술은 활동도가 높을 것으로 기대되고 있지만, 이를 악용할 경우 범죄에 이용될 가능성이 있으며, 죽은 사람의 목소리를 흉내 낸다는 점에서 윤리적인 문제, 고인 음성 데이터에 대한 소유권 분쟁 가능성 등의 문제가 제기되었다.

- AI-05
  - 사고명 : AI 챗봇 '이루다' 의 혐오 발언 및 개인 정보 유출로 인한 서비스 중단
  - 사고 내용 요약 : 이루다는 딥러닝 알고리즘을 통해 20대 여대생으로 설정된 AI 챗봇이며, 학습을 통해 자연스러운 일상 대화를 구현한 것이 특징이다. 이루다는 출시 3주 만에 일부 사용자의 이루다 성희롱, 이루다가 쏟아낸 차별과 혐오 발언, 이루다가 무단학습한 개인정보 유출 등의 사유로 서비스가 중단됐다. 이루다는 문제점을 개선해 이루다 2.0으로 시험서비스를 재개한 상태다.

- AI-06
  - 사고명 : 안면인식 AI 프로그램 오류로 무고한 흑인 남성 체포
  - 사고 내용 요약 : 범죄자를 찾는데 쓰인 얼굴 인식 프로그램이 데이터베이스에서 지목한 범인이 실제 무고한 사람들인 경우가 있었는데 피해자가 모두 흑인이었다. 이로 인해 AI에 의존한 사법 정책이 인종차별을 더 악화시킬 수 있다는 우려가 나왔다. 미 국립표준기술원(NIST)이 2019년 마이크로소프트(MS) 제품을 포함해 189개 안면인식 AI를 분석한 결과, 흑인 및 아시아계에 대한 오류 비율이 백인보다 10~100배 높았다. 또한 AI는 남성에 비해 여성을 잘 식별하지 못했고, 노년의 얼굴을 잘못 인식할 확률도 중년의 10배에 달했다.

- AI-07

- 사고명 : 학습데이터 편향성으로 인한 AI 학력평가 알고리즘의 오류
- 사고 내용 요약 : 영국은 코로나로 대입 시험을 치르지 못하자 대안으로 그동안 실시한 예비 시험과 학교 과제 등을 바탕으로 AI 알고리즘을 이용해 성적을 산정하는 방식을 택했다. 학력평가 알고리즘은 학교의 과거 입시 성적을 반영하도록 하였는데, 그 결과 원데이터의 편향성으로 인하여 알고리즘은 노동자 및 소외계층 학생에게는 지나치게 낮은 점수를 주고, 사립학교 학생에게는 지나치게 높은 점수를 주었다. 학생 가운데 40%가 교사가 예측한 점수보다 낮게 받았고, 이에 수백 명의 학생들이 교육부 청사 앞에 모여 이에 항의하는 사태가 벌어졌다.

• AI-08
- 사고명 : AI 의사 '왓슨(Watson)', 국내 실정과 맞지 않는 의료 AI의 한계
- 사고 내용 요약 : 왓슨은 AI을 의료 분야에 적용한 대표적인 사례로 환자의 증상을 입력하면 과거 수십 년간 확보된 임상 사례는 물론 수십만 페이지에 달하는 전문 자료를 검색해 환자에게 가장 적합한 치료법을 제시한다. 국내 7개의 대학 대학병원이 왓슨을 도입했는데 국내 실정에 맞지 않아 약 70%가 재계약을 하지 않았다. 왓슨은 대상국의 상황을 고려하지 못한다는 단점이 있어 한국인의 특성을 제대로 반영하지 못하는 것은 물론, 한국어의 독특함도 제대로 읽어내지 못해 의사와의 의견 일치율이 떨어졌다. 또한 우리나라 데이터 입력 시 추가 비용이 발생했으며, 의사가 환자 정보 데이터 수기 입력해야 하는 불편함도 한계점으로 지적됐다.

• AI-09
- 사고명 : 음성 딥페이크 기술을 이용한 보이스피싱
- 사고 내용 요약 : 영국의 한 에너지 기업의 CEO가 독일 모기업 CEO의 목소리를 모방한 음성 딥페이스 보이스 피싱에 속아 22만 유로를 송금한 사기 사건이 발생했다. 송금 요청을 받은 CEO는 자신과 통화한 상대방이 목소리 뿐 아니라 독일인의 억양과 발성 패턴까지 상당히 닮아있어 대화하는 사람의 신원을 조금도 의심하지 않았다. 이 사건은 딥페이크 기술을 이용한 첫 번째 사기 사건으로 기록됐다.

- AI-10
  - 사고명 : 日 유명 호텔, 잦은 고장과 오류로 로봇 종업원 해고
  - 사고 내용 요약 : 일본 헨나호텔은 로봇 직원이 일하는 호텔로 인기를 끌었다. 오픈 초기에는 약 80여 대의 로봇만을 도입해 체크인과 수하물 운반을 맡기다 프론트, 객실, 로비 등 곳곳에 배치한 로봇이 243대까지 늘어났다. 그러나 AI 로봇의 오작동, 성능 미비, 프론트 로봇의 체크인 오류, 수하물 운반 로봇 사고 등으로 인해 투숙객들의 불만이 잦았고, 시간이 지나면서 로봇 수리 및 업그레이드 비용이 역시 증가했다. 결국 헨나호텔은 로봇의 잦은 오류와 투숙객의 불만으로 절반을 없애고 결국 사람으로 대체했다.

- AI-11
  - 사고명 : 로봇의 오작동으로 인해 체스 로봇과 대국하던 소년이 당한 신체 부상
  - 사고 내용 요약 : 러시아 모스크바에서 열린 국제 체스포럼의 '인간 대 로봇' 체스 경기에 참여한 7세 소년이 함께 체스를 두던 로봇에 의해 손가락 골절상을 당한 일이 발생했다. 로봇이 소년의 장기말을 빼앗자 소년이 이를 막으려 움직였는데 이때 로봇이 소년의 손가락을 움켜쥐었다. 소년이 손가락을 빼기 위해 안간힘을 썼지만 로봇은 손을 놓지 않았고 네 명 성인의 도움으로 소년은 풀려났다. 대회 주최 측은 소년이 너무 빠르게 반응해 동작한 것이 로봇의 과잉 동작을 낳았다며 소년에게 책임을 넘겼다.

- AI-12
  - 사고명 : 드론 오폭으로 인한 민간인 사망 사고
  - 사고 내용 요약 : 2021년 8월 미군은 카불 시내에서 폭탄을 실은 것으로 추정되는 차량을 지목해 드론으로 폭격했다. 그러나 자살폭탄 테러용 폭발물이 실린 것으로 추정된 차량에는 물통들이 실려 있었고 사망자 역시 IS 조직원이 아니라 미국 구호 단체의 전기기술자와 그 일가족 10명이었다. 2021년 12월 미국방부는 비극적 실수였음을 인정했으나 민간인 10명을 숨지게 한 드론 오폭과 관련해서는 아무도 처벌하지 않는다고 발표했다.

- AI-13
  - 사고명 : 유튜브 추천 알고리즘으로 인한 편향성 강화

- 사고 내용 요약 : 유튜브 추천 알고리즘은 수많은 콘텐츠가 누적되면서 알고리즘을 이용해 사용자 관심에 기반을 둔 추천 서비스를 제공한다. 이 알고리즘으로 인해 이용자 입장에서는 방대한 정보 가운데 필요한 정보만 받게 되며, 결국 이용자는 기존에 관심 있던 정보, 가치관 등에 부합하는 정보만 제공받는다. 그러나 이런 추천 알고리즘은 편향성을 강화시킬 수 있는 악순환의 고리를 형성한다는 점에서 치명적이고 단순히 개인의 편견, 고정관념을 넘어 사회 전체에 악영향을 끼칠 수 있다.

• AI-14
  - 사고명 : 배송 드론 추락사고
  - 사고 내용 요약 : 스위스 우정서비스(Swiss Post)와 미국 벤처기업 '매터넷'(Matternet)은 파트너쉽을 맺고 무인 항공기로 병원 간의 실험실 샘플을 운반하는 운송 서비스를 시작했다. 스위스 우정서비스와 매터넷은 약 1년간 총 3000여회에 달하는 비행을 성공적으로 완수했으나 2019년 1월 25일 GPS 관련 부품에 전력을 공급하는 회로에 문제가 생기면서 드론이 취리히 호수에 추락하는 사고가 발생했다. 이 사고로 드론 배송 서비스는 당분간 중단되었고 드론을 업그레이드 한 후 배송 서비스를 재개했다. 그러나 2019년 5월에 보다 심각한 문제로 드론이 제대로 동작하지 못하는 사고가 발생했고, 이 사고 이후 스위스 우정시비스는 드론 배송 서비스를 무한정 중지시켰다.

• Metaverse-01
  - 사고명 : 미성년자 대상 디지털 성범죄
  - 사고 내용 요약 : 제페토(ZEPETO)는 대표적인 국내 메타버스 서비스로 가상세계에서 아바타를 통한 라이프 로깅서비스를 제공하며 미성년자의 이용률이 높은 편이다. 이 제페토에서 30대 남성이 자신을 20대 초반의 대학생으로 속여 미성년자들을 대상으로 디지털 성범죄를 저질러 구속되었다. 가해자는 피해자들과 지속적인 친분을 쌓고 환심을 산 뒤 자신의 신체 사진을 먼저 찍어 보낸 후 피해자들의 신체 사진을 요구한 온라인 그루밍(심리적 지배)을 통해 성 착취물을 제작하여 보관했고, 메신저를 통해 성적인 대화를 지속하였다. 경찰은 향후 메타버스 공간상 디지털 성범죄에 대해 단속을 강화할 것이라고 밝혔다.

- Metaverse-02
  - 사고명 : 현실과 비슷해지는 디지털 성범죄
  - 사고 내용 요약 : 호라이즌 월드는 VR을 기반으로 하는 가상현실 소셜 공간을 제공한다. 2021년 12월 베타테스트 과정에서 여성 아바타가 남성 아바타들에게 집단 성폭행과 성희롱을 당한 내용이 공개되었으며 사건 후 호라이즌은 내부 검토를 통해 아바타 사이의 거리두기가 도입되었다. 그러나 2022년 5월 한 비영리 단체의 여성 연구원이 여성 아바타로 접속 후 또다시 아바타 간 성폭행을 당한 사실이 공개되었는데, VR기기를 통해 불편한 느낌이 실제 사용자에 전달되어 현실과 비슷한 충격을 받는다고 보고해 더욱 충격을 주었다.

- Metaverse-03
  - 사고명 : 메타버스 공간 내 지적 재산권 침해
  - 사고 내용 요약 : 로블록스가 자사 플랫폼에서 음악 저작권자들에게 비용을 지불하지 않고 음악을 불법으로 사용하고 있다는 이유로 전미음악출판협회에서 로블록스를 상대로 손해배상 청구소송을 제기했다. 전미음악출판협회는 로블록스가 사용자들이 음악을 업로드 할 때 마다 수수료 방식으로 수익을 창출하면서, 저작권 라이선스 계약을 체결하지 않고 음반 제작사와 창작자들에게 저작권료를 지급하지 않고 있다고 소송 경위를 밝혔다. 이 소송전은 비슷한 서비스를 제공하는 메타버스 플랫폼과 저작권자들의 권리관계와 계약 문제를 정리해 줄 것이라는 기대 속에 판결 결과가 주목받고 있었으나, 두 당사자가 분쟁 사항에 대한 합의로 마무리 되었다.

- Metaverse-04
  - 사고명 : 가상부동산 서비스 중단으로 인한 유료 구매자의 금전적 손실
  - 사고 내용 요약 : 세컨서울은 서울 전역을 가상부동산 타일 694만개의 대체불가능토큰으로 나누고 타일 구매자들이 가상부동산을 통해 투자, 임대, 광고 수익을 얻도록 개발된 메타버스 플랫폼이다. 론칭 당시 타일 액면가는 1만원이었고, 사전 신청 인원수가 25만명이 이를 정도로 큰 기대를 모았으나 이틀 만에 서비스 점검을 이유로 서비스가 중단됐다. 세컨서울 사태와 관련하여 모회사가 개발자를 상대로 배임, 금융거래법 위반 등의 소송이 진행 중이다. 또한 모회사는 양 측의 서비스 사용권 및 소송 문제로 기존 세컨서울 서비스는 폐기하

고 메타서울이라는 신규 플랫폼을 오픈 준비 중이다.

- Metaverse-05
  - 사고명 : 가상세계 과몰입으로 겪는 신체 부상
  - 사고 내용 요약 : VR기기를 이용해 가상 세계에 접속하는 이용자가 많아지면서 가상 세계에 지나치게 과몰입 한 이용자가 현실에서 각종 신체 부상을 당하는 사례가 늘고 있다. 메타버스 공간은 현실에서 이용자의 시각과 청각 등 기능을 차단 시키기 때문에, 공간감을 잃게 만들 수 있어 신체 부상의 위험성도 있다. 이에 전문가는 메타버스 이용시 가급적 자주 휴식을 권하고 있고 기업들은 메타버스 이용 시 이용자의 안전을 보장하기 위한 각종 기능을 추가하고 있다.

- Metaverse-06
  - 사고명 : AR 게임 포켓몬 고로 인해 발생한 교통사고 및 안전사고
  - 사고 내용 요약 : 대표적인 위치 기반 AR 게임인 '포켓몬 고(Pokemon Go)'는 포켓몬이 실제 세계에 존재하지 않지만 현실 세계와 영상 콘텐츠의 결합으로 인해 이용자들이 포켓몬스터를 잡아 볼 수 있는 경험을 제공해 큰 인기를 끌었다. 그러나 과몰입하는 사용자가 많아지면서 보행 중, 혹은 운전 중 포켓몬 고 게임을 하는 유저들에 의해 최대 사망에 이르는 교통사고 다수 발생했고, 철도의 선로 위, 원자력 발전소의 일반인 출입 금지 지역, 재해 지역의 피난 명령 지역 등에서도 포켓몬이 출현해 이용자들의 침입으로 인한 안전사고 우려가 커졌다.

- Metaverse-07
  - 사고명 : 메타버스의 플랫폼화를 가로막는 주요 난제 중 하나인 VR 울렁증
  - 사고 내용 요약 : 3D, VR 게임 이용이나 영상 시청 시, 일부 이용자들이 피로감이나 두통, 구토 동반하면서 멀미와 비슷한 육체적 증상을 겪고 있다. 이를 VR 울렁증이라 하는데 주요 원인 중 하나는 지연(레이턴시) 문제다. VR 울렁증이 이슈가 된 초기에는 멀미약을 복용하거나 긴 시간 게임 이용을 자제하고 휴식을 취하는 등의 해결법이 제기 됐지만, VR 기술이 고도화되면서 VR 울렁증을 기술적으로 해결하려는 움직임도 잇따르고 있다.

- Blockchain-01
  - 사고명 : 크로스 체인 브릿지에서 발생한 대규모 해킹
  - 사고 내용 요약 : 액시 인피니티는 액시즈라 불리는 NFT 형태의 디지털 애완 동물을 싸우고 양육하고 팔아서 암호화폐를 버는 게임으로, 액시 인피니티 개발사가 소유한 블록체인 네트워크인 로닌 네트워크를 통해 플레이어들은 액시 인피니트에서 번 디지털 코인을 이더리움과 같은 다른 암호화폐와 교환할 수 있다. 2022년 3월 한 이용자가 보유한 이더리움 출금이 안된다는 신고를 한 후 해킹을 당한 사실 밝혀졌는데 이 해킹은 액시 인피니티와 이더리움 등 다른 블록체인을 연결해주는 이른바 '로닌 브릿지'에서 발생했다. 서로 다른 블록체인 플랫폼 간 자산 이동을 돕는 기술인 크로스 체인 브릿지에서 발생한 것으로 알려지면서 유사한 기술을 사용하고 있는 여러 디파이(DeFi · 탈중앙화금융)의 신뢰성까지 흔들고 있다는 분석이 나왔다.

- Blockchain-02
  - 사고명 : 알고리즘 기반 스테이블 코인의 폭락으로 인한 투자자 손실
  - 사고 내용 요약 : 한국계 블록체인 프로젝트 테라(Terra)에서 사용하는 루나(LUNA) 코인은 변동성이 취약점으로 꼽히는 다른 가상화폐와는 달리, 가치가 실물자산에 연동하도록 설계된 스테이블 코인 방식을 활용해 가격 변동성을 줄인다는 장점을 내세웠다. 알고리즘 기반 스테이블 코인은 담보물이 필요 없는 대신 정교하게 짜인 프로그래밍 코드로 코인의 가치를 일정하게 유지하도록 설계되었다. 그런데 이 알고리즘에 허점이 보이면서 2022년 5월 초만 하더라도 시가총액 300억 달러(약 38조원)를 기록하던 루나가 한순간에 99.99%까지 가격이 하락해 투자자들이 큰 손실을 입으면서 국내를 넘어 전 세계 가상자산 시장을 얼어붙게 하는 악재로 작용했다.

- Blockchain-03
  - 사고명 : NFT 거래소 해킹
  - 사고 내용 요약 : 오픈씨(OpenSea)는 현재 세계 최대의 종합 NFT거래 플랫폼으로 사용자는 플랫폼에서 NFT를 발행, 전시, 거래 및 경매를 할 수 있으며 스마트 계약 기술을 이용해 독특한 NFT를 사고 판매한다. 2022년 2월 19일 오

후 5시부터 8시까지 오픈씨 이용자 32명의 계정에서 비정상적인 NFT 전송 사고가 발생하여 일부 NFT 소유권이 무상으로 양도되었는데, 이로 인해 오픈씨 32명 사용자가 보유한 254개 NFT가 탈취되었으며, 170만달러(약 20억원) 상당의 피해 발생했다. 해킹 수법은 해커들이 32명의 이용자들에게 불법 계약서에 서명을 하게 만드는 방식이었고, 이후 오픈씨는 스마트 계약 표준을 새롭게 업데이트하고 있는 것으로 알려졌다.

- Blockchain-04
  - 사고명 : 블록체인 기반 NFT 신종 사기
  - 사고 내용 요약 : 2021년 11월 25일 ' 캣슬 ' 이라는 명칭을 가진 프로젝트는 희소가치가 있는 고양이 캐릭터로 게임이나 상품을 만들어 투자 가치를 만들겠다는 사업 구상과 함께 매달 가상화폐를 지급하겠다며 투자자를 모았다. 이렇게 홍보해 최초 3만원대에 팔리던 고양이 캐릭터는 최대 50만원까지 가격이 치솟았고 3백명이 넘는 투자자가 모집됐으며, 2억 7천만원 상당의 고양이 캐릭터 5천개가 판매됐다. 이후 프로젝트 운영자들은 3개월도 안돼 잠적했는데, 이 사건은 가상화폐 생태계에서 개발자가 갑자기 프로젝트를 중단하고 투자금을 들고 사라지는 러그풀이란 신종 사기 수법을 말한다. 한편 프로젝트 운영자들은 2022년 4월 검거되었다.

- Blockchain-05
  - 사고명 : 솔라나 네트워크 장애
  - 사고 내용 요약 : 솔라나는 저렴한 거래 수수료와 빠른 전송 속도를 무기로 최근 급부상한 블록체인 플랫폼이다. 그러나 많은 거래량이 몰리거나 낮은 수수료로 인해 오히려 분산서비스거부(DDoS) 공격이 잦아지면 블록 생성 오류로 암호화폐 거래 장애 및 해당 플랫폼을 기반으로 하는 서비스들의 네트워크 장애가 많이 발생하고 있다. 솔라나를 향한 디도스 공격은 2021년에만 세 차례 이상 이루어졌다.

제3절 언론기사 빅데이터 분석을 통한 SW 신기술 안전 사고 조사

1. 언론기사 빅데이터 분석 개요

정성 연구의 SW 신기술 안전 사고 조사의 전문성을 확보할 수 있는 방안으로 본 연구에서는 언론기사 빅데이터 분석을 활용했다. 기간은 2017년 1월 1일 부터 2022년 6월 30일까지의 66개월로 선정했으며, AI, 메타버스, 블록체인과 관련 언론기사를 선별하여 수집했다. 언론기사의 언어는 글로벌 트렌드를 고려하여 영어로 선정했다.

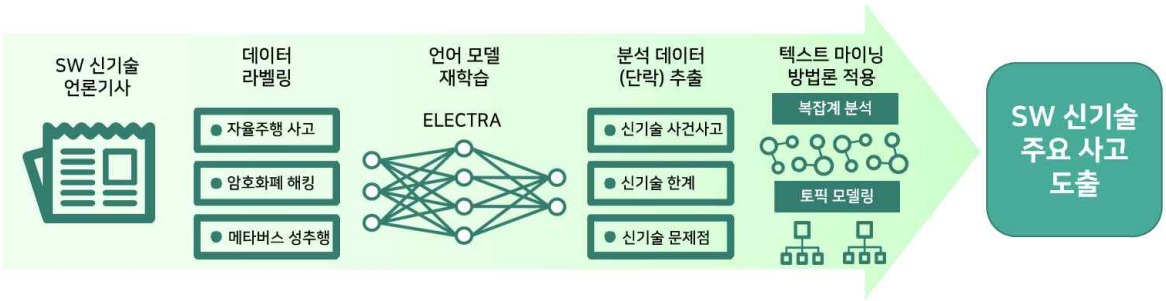
언론기사 빅데이터 분석의 가장 큰 어려움은 수십만 건의 SW 신기술 관련 언론기사에서 사건이나 사고를 다룬 언론기사를 선별하는 것이다. 이는 기본적으로 문서 분류의 과업에 해당하나, 언론 기사의 특성상 문서 분류의 효용을 기대하기는 어렵다. 그 이유는 다양한 예외가 존재하기 때문이다. 예를 들면, 사건이나 사고를 다루고 있으나 원인이 SW 신기술이 아닌 경우, 중립적인 사실 전달과 배경 설명으로 인해 중요한 데이터의 잠재적인 제외가 일어나는 경우가 있을 수 있다.

이에 본 연구에서는 분석 대상 데이터를 언론 기사에서 언론 기사를 구성하는 단락으로 선정했다. 이는 중립적인 사실 전달과 배경 지식의 설명을 제외함에 있다. 그렇다면 언론기사 빅데이터 분석의 데이터는 “SW 신기술과 관련된 사건 사고 혹은 한계 및 문제점을 기술하고 있는 단락”으로 이해할 수 있다. 이것은 분석하고자 하는 대상과 그렇지 않은 것을 이진 분류하는 과업을 해결함으로써 데이터를 확보할 수 있을 것이다. 이를 위해 본 연구에서는 언어 모델 재학습을 활용했다.

언어 모델 재학습을 통해 도출된 “SW 신기술과 관련된 사건 사고 혹은 한계 및 문제점을 기술하고 있는 단락”을 “분석 대상 데이터”로 정의하고, 이를 분석하기 위해 텍스트 마이닝 기법을 활용했다. 텍스트 마이닝은 크게 두 가지 접근을 활용했는데 첫 번째는 토픽 모델링(Topic Modeling)이다. 토픽 모델링은 다수의 문서 혹은 단락을 유사한 주제로 군집해주는 비지도학습 기법의 일종이다. 토픽의 수는 초모수의 하나로 토픽 모델링 결과의 응집성으로 도출하거나, 연구자의 직관에 의해 결정하기도 한다. 특정 토픽은 그것을 대표하는 키워드가 나타나며, 키워드를 조합하여 토픽의 주제를 정성적으로 결정한다. 본 연구에서는 토픽 모델링을 SW 신기술의 사고 분야를 구분하는데 활용하였다. 또한 토픽에 해당하는 단락의 수에 대한 연도별 변화는 분야별 사고의 추이를 확인할 수 있다는 점에서 매우 유용한 결과를 도출할 수 있다.

두 번째 텍스트 마이닝 기법은 키워드 네트워크 분석(Keyword Network Analysis)이다. 키워드 네트워크 분석은 네트워크 분석의 일종으로 특히 문서 단위로 출현하는 단어 간의 관계를 통해 현상을 분석하는 기법이다. 키워드 네트워크 분석에서는 키워드 네트워크를 정의하는 방식이 여러 가지 있을 수 있으나, 언론기사의 특성을 반영해 연결된 단어가 얼마나 많은 문서에 등장하는지로 설정했다. 여기에 키워드 네트워크를 구성하는 데이터는 토픽 모델링의 각 토픽에 해당하는 단락이다. 이후 키워드 네트워크의 군집 분석의 일종인 커뮤니티 분석을 수행하여 SW 신기술의 주요 사고를 도출한다.

언론기사 빅데이터 분석의 전반적인 흐름은 [그림 2-2]와 같다.



[그림 2-2] 언론기사 빅데이터 분석의 흐름

[그림 2-2]에서는 3대 SW 신기술이 모두 포함되어 있으나, 분야별로 특성이 다르기 때문에 3대 분야를 나누어서 분석을 실시했다. 그러나 메타버스의 경우 메타버스의 사회 경제적 시스템을 블록체인으로 구현하는 접근이 두드러지며, 메타버스와 관련된 용어집에도 대부분 블록체인 기술을 포함하고 있다. 이에 본 연구에서는 AI와 메타버스/블록체인의 두 가지 데이터를 각각 분석하는 접근을 취했다.

2. 언론기사 빅데이터 수집 및 전처리

본 연구에서는 언론기사 수집을 위해 영국 다우 존스에서 운영하는 팩티바(Factiva) 데이터 베이스를 활용했다. 팩티바는 전 세계 3.3만 여개 언론사에서 등록하는 모든 언론기사를 수집하여 서비스하는 플랫폼이며, 각 언론기사 마다 고유 부가정보(attribute)를 부여하여 원하는 데이터를 용이하게 확보할 수 있다. 부가정보 중 유용한 것은 특정 산업 분야를 지칭하는 산업 코드(Industry code)와 언론기사의 주제를 분류하는 주제 코드(Subject code)가 있다. 산업 코드와 주제 코드만으로도 본 연구에서 수

집하고자 하는 AI, 메타버스, 블록체인을 어느 정도 구분할 수 있으나, 팩티바에서 제공하는 코드가 다소 광범위하게 설정되어 있어 부여된 코드와 실제 언론기사가 상이한 경우가 존재한다.

팩티바는 다국어를 지원한다는 점에서 언어를 선택할 수 있는데, 본 연구에서는 글로벌 차원의 SW 신기술의 사건 사고를 다룬다는 점에서 영어를 선택했다. 팩티바에서는 또한 언론사 목록을 제공하는데, 대륙별 주요 언론사 목록, 영미권 주요 언론사 목록 등 영향력있는 언론사를 선별할 수 있다는 장점이 있다. 본 연구에서 선택한 3개 언론사 목록은 다음 <표 2-1>과 같다.

<표 2-1> 언론기사 수집을 위한 팩티바 언론사 목록

언론사 목록	설 명
Major News & Business Sources	- 글로벌 뉴스와 비즈니스 주요 언론사 목록 - 9개(아프리카, 아시아 태평양, 호주와 뉴질랜드, 캐나다, 유럽, 라틴아메리카, 중동, 영국, 미국)의 지역으로 구분
Top Industry	- 산업 전반에 걸쳐 언론기사를 생산하는 권위있고 영향력있는 언론사 목록
Top US/UK Sources	- 미국과 영국의 주요 언론사와 지역지를 포함

자료 : Factiva, Dow Jones

<표 2-1>에 해당하는 언론사 중 영어 기사를 제공하는 언론사는 총 331개이다.



[그림 2-3] 언론기사를 수집하는 331개 언론사의 분포

[그림 2-3]은 331개 언론사의 분포를 나타내며, 우리나라도 7개 언론사[2]가 포함되어 있다. 언론기사 수집기간은 2017년 1월 1일 부터 2022년 6월 30일까지 총 66개월간으로 구성되며, 전술한 바와 같이 AI와 메타버스/블록체인의 두 가지 영역으로 나누어 수집을 추진했다. 데이터 수집 조건은 다음 <표 2-2>와 같으며, 도합 약 35만 건의 SW 신기술 관련 언론기사를 수집했다.

<표 2-2> 언론기사 데이터 수집 조건

기 준		AI	메타버스/블록체인
기간		2017.01.01. ~ 2022.06.30. (66개월)	
분야주)	산업코드	Autonomous Driving, Connected Vehicle, Artificial Inteliigence, Software Agent	Blockchain, Virtual Reality
	주제코드	Artificial Intelligence / Machine Learning, Automation, Robotics	-
검색 키워드 / 검색 방식		<표 2-3> 참고 / 핵심 키워드가 헤드라인에 검색, 활용 키워드가 본문에 2회 이상 검색	<표 2-4> 참고 / 검색 키워드가 본문에 2회 이상 검색
수집된 기사 건수 (단락 수)		221,704 건 (3,955,439 개)	132,511 건 (2,784,365 개)

주) 분야는 팩티바에서 직접 언론기사별로 제공하는 부가정보이나, 정확성을 담보하지는 않음

<표 2-2>의 데이터 수집 조건을 보면 팩티바에서 제공하는 부가정보인 산업 코드와 주제 코드 이외에 별도의 검색 키워드를 활용했다. 그 이유는 분야별로 SW 신기술을 내포하는 고유명사를 사용하기 때문이다. 이를 위해 별도의 검색 키워드를 선별하고, 각각의 분야에 맞는 검색 방식을 선택했다.

<표 2-3> AI 분야 검색 키워드

기 준	키워드 목록
핵심 키워드	artificial intelligence, deep learning, machine learning, robotics, intelligent agent
활용 키워드	digital footprint, marketing automation, targeted advertise, machine translation, face recognition, face detection, game bot, AI for social good, precision agriculture, digital agriculture, intelligence tutoring system, algorithmic trading, automated trading system, medical image analysis, lawbots, virtual assistant, voice assistant, chatbot, intelligent personal assistant, deepfake, video content analysis, computer-generated news, automated journalism, self driving, autonomous car, autonomous vehicle, driverless car

2) 가나다 순으로 동아일보, 매일경제, 연합뉴스, 조선일보, 중앙일보, 코리아해럴드, 한국일보

먼저 AI의 경우는 <표 2-3>과 같이 핵심 키워드와 활용 키워드로 구분하였다. 핵심 키워드는 AI를 대표하는 키워드로 제목이나 헤드라인에 한 번이라도 검색된 경우 AI와 관련된 기사로 보았다. 또한 활용 키워드의 경우에는 하나의 기사에서 특정한 하나의 활용 키워드가 2회 이상 검색되었을 때를 AI 관련 기사로 보고 수집 대상에 포함시켰다. 활용 키워드 목록은 위키백과의 "AI 활용 분야"에서 고유명사를 특정했다.

<표 2-4> 메타버스/블록체인 분야 검색 키워드

기 준	키워드 목록
메타버스 분야 검색 키워드	metaverse, digital twin, volumetric, augmented reality, virtual reality, extended reality, head mounted display, mirror world, lifelogging
블록체인 분야 검색 키워드	blockchain, web3, non-fungible token(nft), play-to-earn(p2e), decentralized autonomous organization(dao), decentralized application(dapp), decentralized finance(defi), decentralized exchange ※ 약어의 경우도 검색 키워드로 활용

메타버스/블록체인 분야는 <표 2-4>의 검색 키워드가 하나의 언론 기사에 특정 키워드가 2회 이상 검색될 경우를 포함한다. 검색 키워드는 메타버스와 블록체인을 통합한 계기가 되었는데, 메타버스의 가상 사회를 지탱하는 경제 체계를 블록체인으로 구현한 경우가 많았기 때문이다. 또한 블록체인 분야에서는 암호화폐와 관련된 키워드를 배제했는데, 그 이유는 암호화폐의 등락을 고지하는 수많은 단발성 기사는 본 연구에서 수집하고자 하는 언론기사가 아니기 때문이다.

이러한 검색 키워드를 조건으로 언론기사를 수집한 결과 AI 분야는 221,704건의 언론기사가 수집되었으며, 본 연구의 분석 단위 데이터인 단락의 수는 3,955,439개 이다. 메타버스/블록체인의 경우에는 총 132,511건의 언론기사가 수집되었으며, 이는 2,784,365개의 단락으로 구성되어 있다.

이후에는 기본적인 데이터 전처리를 수행했다. 본 연구의 분석 대상 데이터는 SW 신기술과 관련된 사건 사고, 문제점 및 한계를 언급하는 단락이기 때문에, 언론기사를 다수의 단락으로 분리하는 작업을 거쳤다. 팩티바에서 수집한 데이터는 언론기사 자체의 제목과 내용 이외에 다양한 부가정보가 부여되어 있다. 본 연구에서는 [그림 2-4]와 같이 수집된 언론기사 데이터에서 <표 2-5>와 같은 데이터 프레임을 설정하여 데이터의 일차적인 전처리를 수행했다.

```
{
    "copyright":"Copyright © 2020 Yicai Global Co., Ltd. All Rights Reserved. ",
    "subject_codes":",ccat,",
    "art":"",
    "modification_datetime":"1594172606000",
    "body":"The platform can also randomly throw in extreme environmental variables. It may take on
e month to complete an extreme test case on actual roads, while the platform can add rain, snow and
poor lighting at night, in just 30 seconds using AI. It can construct millions of scenarios a da
y.\n\n
    Conventional platforms lack data on extreme scenarios and cannot restore the uncertainties of r
eal road conditions, so automatic driving training is less efficient and is unlikely to achieve tec
hnical breakthroughs. The hybrid simulation test platform can solve this problem and push the techn
ology to the highest level of 5, meaning fully autonomous driving without human intervention, DAMO
added.\n\n
    Alibaba announced its formal entry into the autonomous driving sector in 2018, with DAMO Academ
y developing a self-driving platform AutoDrive. The firm had earlier established a self-driving joi
nt venture Banma Network Technology with SAIC Motor in 2015, with its products applied to a variety
of production models.\n\n
Editor: Peter Thomas\n\n
    Click here to view image.",
    "company_codes_occur_ticker_exchange":"",
    "company_codes_occur":",shautm,",
    "company_codes_about":",alibab,",
    "company_codes_lineage":"",
    "company_codes_ticker_exchange":"",
    "snippet":"(Yicai Global) April 22 -- DAMO Academy, Alibaba’s research institute, today unveile
d the world’s first hybrid test platform for autonomous driving simulations that can virtually test
eight million kilometers a day, greatly boosting the efficiency of artificial intelligence training
for self-driving, The Paper reported.\n\n
    The platform combines virtual and real simulation technologies that use real drive test data to
automatically generate mockup scenarios, and imitate real-time situations such as acceleration, sha
rp turns and emergency stops through random artificial intervention, DAMO said.",
    "company_codes_relevance_ticker_exchange":"",
    "publication_date":"1587513600000",
    "market_index_codes":"",
    "credit":"",
    "section":"Tech",
    "company_codes_association_ticker_exchange":"",
    "currency_codes":"",
    "company_codes_about_ticker_exchange":"",
    "region_of_origin":"ASIA CHINA CHINAZ EASIA ",
    "company_codes_lineage_ticker_exchange":"",
    "ingestion_datetime":"1594172606000",
    "modification_date":"1623963233229",
    "source_name":"Yicai Global",
    "language_code":"en",
    "region_codes":"",
    "company_codes_association":"",
    "person_codes":"",
    "byline":"Dou Shicong ",
    "company_codes_relevance":",shautm,",
    "source_code":"YICAIG",
    "an":"YICAIG0020200708eg4m000yd",
    "word_count":"264",
    "company_codes":",shautm,shautm,alibab,",
    "industry_codes":",i64,i656000301,iecom,iint,iretail,itech,",
    "title":"Alibaba’s DAMO Academy Unveils First Hybrid Test Simulator for Self-Driving",
    "publication_datetime":"1587513600000",
    "publisher_name":"Yicai Global Co., Ltd",
    "action":"add",
    "document_type":"article"
}
```


[그림 2-4] 팩티바에서 수집된 언론기사의 예시

<표 2-5> 데이터 프레임

구분	num	publication _datetime	snippet	body	num _paragraph	source _name	subject _code	industry _code
내용	고유번호	발간 시간	헤드라인 / 요약 단락	본문내용 (단락구분)	단락의 수 (snippet + body)	언론사 명	주제 코드	산업 코드
데이터 타입	정수	정수 (long)	문자열	리스트 (문자열)	정수	문자열	리스트 (문자열)	리스트 (문자열)

데이터 프레임으로 전처리를 거친 데이터는 이후 기본적인 예외 처리를 진행했다. 예외 처리는 특정할 수 있는 자동화 수단이 전무하기 때문에, 직접 데이터를 보고 확인하며 제외할 기사나 단락을 선별하는 과정으로 진행된다. 특히 데이터의 전반을 살펴보면서 분석 대상 데이터와 관련이 없는 것은 과감히 배제하였다. 데이터 예외 처리 과정은 데이터 프레임에서 단락을 랜덤 샘플링 하여 예외 처리를 할지의 여부를 일일이 확인한 결과 다음 사항에 대해서 예외처리를 수행했다.

- 단락의 글자 수가 10자 미만, 3,000자 초과일 경우
- 주가, 선물 등 금융상품의 등락을 단순 전달
- 특허 관련 기사 및 특허 문서 자체
- 키워드를 단순히 나열하는 단락
- 연구 결과나 다른 언론기사를 단순 인용했을 때

이러한 기준으로 예외처리를 한 결과는 <표 2-6>과 같다. 여기서 본 연구의 분석 단위 데이터는 단락이라는 점과 언론기사의 특성을 복합적으로 고려해 볼 때, 대다수의 단락은 중립적인 사실 전달과 배경지식에 관한 내용을 차지할 것이다. 따라서 예외처리를 완료한 단락에서 AI는 <표 2-3>의 키워드가 메타버스/블록체인은 <표 2-4>의 키워드가 한 번이라도 등장한 경우(중복 제외)를 분석 데이터로 삼았다.

<표 2-6> 예외처리를 통한 분석 데이터 현황

구 분	예외처리 전		예외처리 후		분석 데이터	
	AI	메타버스/블록체인	AI	메타버스/블록체인	AI	메타버스/블록체인
기사수	221,704	132,511	206,220	121,413	-	-
단락수	3,955,439	2,784,365	3,400,503	2,348,709	373,760	334,083

분석 데이터는 66개월간 수집된 AI, 메타버스, 블록체인 분야의 35만 여건 언론 기사에서 추출된 70만 여개의 단락으로, 이것은 각 분야별 키워드가 실제로 등장한 단락이기 때문에 SW 신기술을 직접적으로 언급하고 있는 언론기사 단락이라고 해석할 수 있다.

3. 언어모델을 활용한 언론기사 빅데이터 정제

이제 남은 것은 분석 데이터에서 SW 신기술과 관련된 사건 사고, 한계 및 문제점을 언급하는 단락(이하 “분석 대상 데이터”로 표현)을 찾는 것이다. 문제를 더 단순화하면 분석 데이터를 구성하는 특정 단락이 분석 대상 데이터인지 여부를 판단하는 이진 분류이다.

문제 자체는 간단하고 쉬워보이지만, 분석 대상 데이터를 인식할 수 있는 분류기를 만들어야 한다는 점에서는 매우 어려운 과업이라 볼 수 있다. 그 이유는 언론기사 단락의 다양성을 모두 고려한 분류기를 만들 수 있어야 하기 때문이다.

본 연구에서는 최근 AI 분야에서 각광을 받고 있는 사전학습 언어 모델(Pretrained Language Model)을 활용했다. 여기서 언어 모델은 주어진 단어나 문장 다음에 올 단어를 예측하는 것으로, 사람이 구사하는 언어 자체를 모델화시킨 것으로 이해할 수 있다. 언어 모델은 AI의 자연어 처리(Natural Language Processing) 분야에서도 가장 어려운 과업 중의 하나로, 완벽한 언어 모델은 거의 모든 자연어 처리 세부 과업(downstream task)들을 해결할 수 있다.

사전학습 언어 모델의 트렌드는 2018년 구글과 OpenAI에서 공개한 BERT(Bidirectional Encoder Representation Transformer)와 GPT(Generative Pretrained Transformer)가 대표적이다. 이 두 가지 모델은 모두 트랜스포머(Transformer)를 활용한 언어모델로 인공신경망이 주류였던 딥러닝의 트렌드를 트랜스포머로 전환시키는 주요한 역할을 했다.

이 사전학습 언어 모델의 장점은 크게 두 가지로 요약할 수 있다. 첫 번째는 다양한 자연어 처리에 적용이 가능한 범용성을 가졌다는 점이다. 이는 본 연구에서 목표로 하는 언론기사 단락의 이진분류에도 적용될 수 있다. 두 번째는 특정 자연어 처리 과업을 수행하기 위해 수천건 수준의 데이터를 학습시키는 것만으로도 우수한 성능을 달성할 수 있다는 점이다.

따라서 사전학습 언어모델을 활용하여 분석 대상 데이터를 이진분류하기 위해서는 가장 먼저 해야할 것이 학습 데이터를 구축하는 것이다. 본 연구에서의 분석 대상 데이터는 SW 신기술의 사건 사고, 한계 및 문제점을 언급하고 있다는 점에서 모두 부정적인 내용을 다루고 있다. 즉, 부정적인 내용을 담고 있는 단락을 일차적으로 선별한 뒤 그것이 실제로 본 연구에서 분류하고자 하는 분석 대상 데이터인지 여부를 라벨링하여 학습 데이터를 구축할 수 있다.

특정 단락이 부정적인지 여부를 판별하는 것은 일종의 감정 분석(Sentiment Analysis)으로 이해할 수 있다. 감정 분석은 최근 마케팅 수단으로 급부상하고 있어 제품 출시 후 SNS에서 일어나는 반응을 분석하는데 활용되고 있다. 가장 용이한 감정 분석 방법은 규칙 기반 방법론이다. 규칙 기반 방법론은 특정 어휘별로 부여된 감정 점수를 문맥에 맞게 조합하여 산출하는 것으로 계산의 부담이 적은 특성이 있다. 그러나 다의어에 대한 문맥적 해석은 어렵다는 점[3]이 단점이라고 볼 수 있다.

이에 본 연구에서는 <표 2-5>와 같이 분석 데이터 70만 건에 대해 규칙 기반 감성 분석 도구 중 하나인 Vader(Valence Aware Dictionary and sEntiment Reasoner)를 적용했다(Hutto and Gilbert, 2014). Vader는 기본적으로 문장 단위로 감성 점수를 부여하는데, 단락의 경우도 적용가능하다. Vader 감성 점수는 일차적으로 긍정, 부정, 중립일 확률[4]을 산출하고, 이 세 확률을 조합한 합성 점수(compound)도 산출한다. 합성 점수는 -1부터 1 사이의 값으로 부여되며, -1에 가까울수록 부정, 1에 가까울수록 긍정이며 0 근처는 중립으로 해석할 수 있다. 합성 점수를 별도의 지표로 활용하는 이유는 대부분의 단어가 중립적이기 때문에 긍정과 부정의 구분을 조금 더 확연하게 나타내기 위함에 있다. Vader에서는 합성점수를 기준으로 0.05 이상을 긍정, -0.05 이하를 부정으로, 나머지는 중립으로 보는 것을 추천한다.

본 연구에서는 학습 데이터를 구축하기 위하여 우선적으로 70만 여개의 분석 데이터에 Vader를 적용했다. Vader의 결과값이 가장 부정적인 단락부터 순차적으로 라벨링을 수행했다. 그 이유는 부정적인 단락에서 분석 대상 데이터를 선별하기 위해서이다. 예를 들면 특정 단락은 부정적인 내용으로 전개되나 그것을 SW 신기술로 해결할 수 있다는 사실을 다루고 있다. 이것은 SW 신기술의 긍정적인 영향을 표현한 것으로 이러한 단락은 분석 대상 데이터에서 제거해야 한다.

3) “block”이라는 단어의 경우 어떠한 시스템을 구성요소로도 표현되는 반면, 규칙 기반 감성 분석 도구에서는 “가로막다”라는 부정적인 단어로 활용된다.
4) 세 확률의 합은 1이다.

학습 데이터의 양은 많을수록 좋지만 시간을 무한정 투입할 수 없으므로, 가용한 시간과 자원을 활용하여 구축해야 한다. 언어모델 재학습에 필요한 데이터의 양은 경우에 따라 다르지만 보통 수천건 수준으로도 어느 정도의 성능을 달성할 수 있다. 이것은 어디까지나 경험적인 영역이기 때문에 양에 대한 절대 수치는 없으며, 학습된 결과가 만족스럽지 않다면 더 많은 학습 데이터를 구축하는 순환 구조로 생각하면 된다.

본 연구에서는 <표 2-7>과 같이 AI 분야 분석 데이터로 2,282건, 메타버스/블록체인 분야 분석 데이터로 2,000건을 학습데이터로 구축했다.

<표 2-7> 언어모델 재학습을 위한 학습 데이터

구 분	AI	블록체인 / 메타버스
데이터 수	2,282	2,000
분석 대상 데이터인 경우주)	1,141	928
분석 대상 데이터가 아닌 경우	1,141	1,072

주) SW 신기술과 관련된 단락이 SW 신기술로 인한 사건 사고, 한계 및 문제점을 언급하고 있는 경우

학습 데이터 중 분석 대상 데이터의 여부는 1:1에 가깝게 산출했다. 이는 Vader를 통해 산출된 부정적인 단락이라고 하여도 그 세부적인 내용이 분석 대상 데이터와 관련이 없는 경우가 많았다. 또한 분석 데이터인 70만여개의 단락에서 분석 대상 데이터의 비율이 어느정도 되는지 가늠을 할 수 없기 때문에, 학습 성능을 위해 분류별로 동일한 데이터를 확보할 수 있도록 고안하였다.

이제 구축된 데이터를 언어 모델에 재학습 시키기 위해 사전학습 언어 모델을 선정하는 과정을 거쳤다. 사전학습 언어모델은 다양한 모델이 오픈소스로 활용가능하며, 이를 쉽게 활용할 수 있는 인터페이스로는 허깅페이스(huggingface) 저장소가 있다. 허깅페이스에서는 다양한 종류의 사전학습 언어 모델을 제공하고 이를 재학습 시킬수 있는 방법을 제안하고 있어 손쉽게 활용할 수 있다는 장점이 있다.

본 연구의 과업은 문단의 이진 분류이다. 문단의 분류에 좋은 성과를 나타내는 모델은 2020년 구글이 공개한 ELECTRA(Efficiently Learning an Encoder that Classifies Token Replacements Accurately)와 2021년 마이크로소프트가 공개한 DeBERTaV3 (Decoding-enhanced BERT with disentangled attention Version 3)이 있다. ELECTRA는 BERT의 사전학습 언어 모델 방식인 마스크드 언어 모델(Masked Languege Model)을

더 효율적으로 접근하기 위해 GAN(Generative Adversarial Network)의 방식으로 변경된 토큰을 탐지(Replaced Token Detection)하는 기법을 적용하여 우수한 성능을 달성했다. DeBERTa의 경우에는 토큰의 위치가 갖는 의미를 부여하기 위해 토큰과 위치의 어텐션을 추가하는 접근을 취했다. 여기서 DeBERTa V3는 사전학습 언어 모델의 방식을 마스크드 언어 모델에서 ELECTRA의 기법인 변경된 토큰 탐지로 변형한 것으로 상당한 성능을 확보할 수 있었다.

본 연구에서는 ELECTRA를 활용하여 재학습을 진행하였다. ELECTRA는 크기에 따라 Small, Base, Large로 구분되는데, 크기가 클수록 더 좋은 성능을 기대할 수 있으나 그만큼 재학습에 소요되는 시간이 길다. 특히 초모수[5]를 변경시켜가며 반복적인 학습을 통해 최선의 모델을 찾아야하는 방법론의 특성상 모델의 크기는 상황에 맞춰 적절한 선택을 하는 것이 필요하다. 재학습은 <표 2-7>의 학습 데이터를 무작위로 학습용과 검증용으로 구분한 뒤 학습용 데이터로 학습을 시키고 검증용 데이터로 추론하여 성능을 측정하는 방식으로 진행된다. 본 연구에서는 학습의 효율을 높이기 위해 NVIDIA RTX A6000 2장으로 구성된 고성능 PC를 활용했다.

이처럼 사전학습 언어 모델을 활용한 자연어 처리 과업의 수행은 꽤 많은 변수가 존재하기 때문에 그 과정과 의미를 정확하게 파악하기 어렵다. 가장 좋은 방법은 재학습 데이터로 다시 학습된 모델에 분석 데이터를 추론시킨 결과가 얼마나 좋은지를 샘플링하여 추정하는 것이다. 본 연구에서는 다양한 ELECTRA 모델과 학습-검증 데이터의 비중 등을 변경시켜 가며 학습시킨 결과[6] <표 2-8>과 같은 재학습 모델을 도출했다.

<표 2-8> 재학습된 언어모델 결과

구 분	AI	블록체인 / 메타버스
모 델	ELECTRA Small	ELECTRA Base
학습 : 검증 데이터 비중	9 : 1	7 : 3
정확도주1)	0.825	0.795
분석 데이터 추론 결과주2)	39,875	10,762

주1) 정확도(Accuracy)는 전체 데이터에서 예측을 올바르게 한 비율을 나타냄

주2) 본 연구에서 “분석 대상 데이터”를 의미함

5) 초모수(Hyper-parameter)는 연구자가 임의로 정할 수 있는 모수로, 학습의 폭을 정하는 학습율, 한꺼번에 학습할 데이터의 양, 학습을 수행하는 최적화 알고리즘 등이 있다.

6) 나머지 초모수는 ELECTRA 논문에서 명시하고 있는 미세 조정(fine-tuning) 초모수를 차용했다.

재학습이 완료된 모델은 AI의 경우 분석 데이터 373,760개 단락에서 AI 관련 사건이나 사고, 한계와 문제점을 언급하는 단락이 39,875개로 산출했다. 메타버스/블록체인 분야의 재학습 모델의 경우 분석 데이터 334,083개 단락에서 메타버스와 블록체인으로 인한 사건 사고 등을 언급하는 단락이 10,762개로 나타났다.

그러나 언어모델 재학습을 통해 도출한 AI 분야 분석 대상 데이터 39,875개 단락과 메타버스/블록체인 분야 분석 대상 데이터 10,762개 단락이 실제로 모든 것을 추출했는지는 확신할 수 없다. 그렇게 하기 위해서는 70만 여개의 분석 데이터를 모두 라벨링해야 하는데, 그렇게 하는 것이 비용적으로 가능하다면 언어 모델 재학습을 할 필요가 없기 때문이다.

추출된 분석 대상 데이터의 적절성을 간접적으로 확인할 수 있는 방법은 있다. 그것은 랜덤으로 샘플링하여 실제로 분석 대상 데이터의 영역에 속하는지를 판단하는 것이다. 본 연구에서는 추출된 분석 대상 데이터에서 각각 1,000건 씩 랜덤 샘플링하여 실제로 분석 대상 데이터인지의 여부를 확인해본 결과 AI 분야에서는 812건이, 메타버스/블록체인 분야에서는 765건이 적절하게 분류되었다.

본 연구에서는 <표 2-8>로 도출된 분석 대상 데이터를 바탕으로 궁극적인 목표인 SW 신기술 안전 사고를 도출하고자 한다. 비록 분석 대상 데이터가 완벽하다고는 보기 어려울 수 있으나 35만여건이나 되는 방대한 언론기사에서 다양한 전처리와 정제 기법을 활용해 분석 대상 데이터를 합리적으로 도출하기 위해 최선의 노력을 다했다.

4. 텍스트 마이닝을 활용한 언론기사 빅데이터 분석

AI, 메타버스, 블록체인과 관련된 지난 66개월 간의 글로별 영문 언론기사 35만 여건에서 SW 신기술로 인해 발생한 주요 사건 사고를 추출하기 위해 SW 신기술의 사건 사고, 한계 및 문제점을 언급하는 단락(“분석 대상 데이터” 로 지칭)을 선별하는 과정을 거쳤다.

기본적인 예외처리를 통한 전처리, 언어 모델 재학습을 통한 단락의 분류 등을 통해 언론기사 35만 여건을 구성하는 약 673만 여개의 단락에서 분석 대상 데이터로 5만 여개의 단락을 추출했다. 이 분석 대상 데이터는 SW 신기술로 인해 발생한 주요 사건 및 사고를 도출하는데 활용하기 위해서 텍스트 마이닝 기법을 활용했다.

본 연구에서 활용한 텍스트 마이닝 기법은 전술한대로 크게 두 가지를 활용한다. 첫 번째는 분석 대상 데이터의 도메인이나 분야를 구분하는 토픽 모델링이고, 두 번째는 도출된 토픽에 대해 키워드 네트워크를 구성하여 다시 군집하는 것이다. 여기서 도출되는 키워드 네트워크 군집을 바탕으로 주요 사건을 추정하여 SW 신기술 안전 사고 양식에 따라 정리했다.

토픽 모델링과 키워드 네트워크 분석은 사이람 넷마이너 4를 활용하여 진행하였다. 전술한대로 두 가지 방법은 토픽 모델링이후 키워드 네트워크 분석을 하는 절차로 이루어져 있기 때문에, 토픽 모델링에 적용된 데이터가 그대로 키워드 네트워크 분석에 활용된다. 본 연구에서 도출한 분석 대상 데이터는 넷마이너에서 제공하는 형태소 분석기가 적용되어 명사만을 추출하게 된다. 그러나 명사에는 지나치게 범용적인 단어[7]가 많기 때문에 적절한 수준에서 제거하는 과정을 거쳤다. 텍스트 마이닝 적용 결과는 방법론의 세부적인 기술보다 결과를 위주로 설명하고자 한다.

먼저 AI 분야의 분석 대상 데이터의 워드 클라우드는 [그림 2-5]와 같다.



[그림 2-5] AI 분야 분석 대상 데이터의 워드 클라우드

워드 클라우드에서 확인할 수 있는 단편적인 키워드만으로도 본 연구에서 도출한 분석 대상 데이터가 대부분 AI와 관련된 사건 사고와 깊은 연관이 있음을 확인할 수 있다.

7) 예를 들면, technology, year, world, use, way, sale, time 등이 있다.

AI 분야 분석 대상 데이터의 토픽 모델링 결과 각 토픽별 상위 키워드를 고려했을 때 총 7개의 토픽이 도출됐다. 7개의 토픽은 본 연구에서 의도한 바와 같이 4개의 토픽은 사건 사고, 3개의 토픽은 한계 및 문제점으로 나타났다. 각 토픽별 키워드와 확률은 다음 <표 2-9>과 같다.

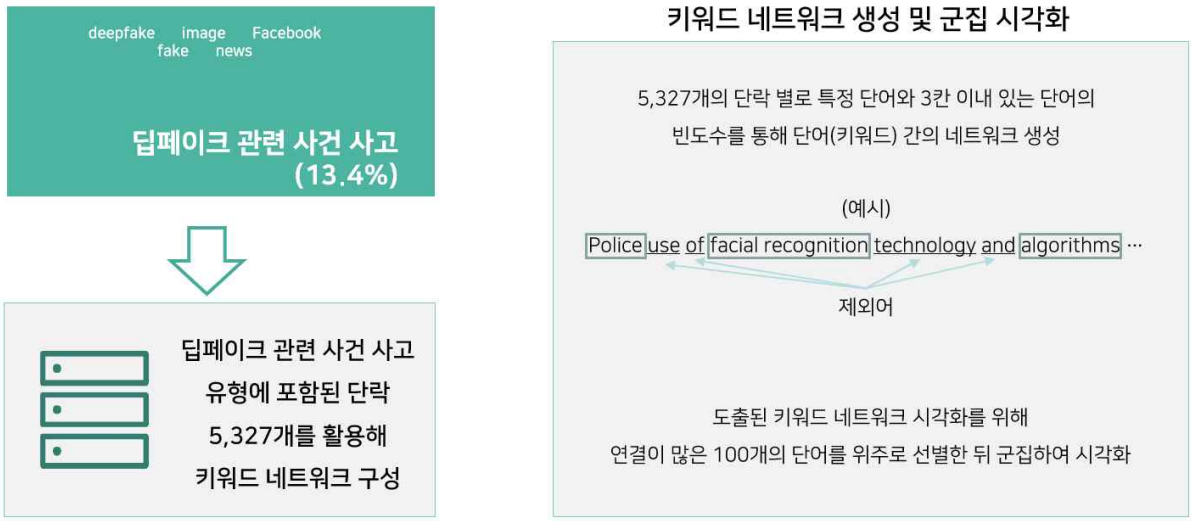
<표 2-9> AI 분석 대상 데이터 토픽 모델링 결과 - 7대 토픽별 키워드

	순위	자율주행 관련 사고 (22.5%)		안면인식 오남용과 감시체계 (20.5%)		딥페이크 관련 사건 사고 (13.4%)		알고리즘과 챗봇의 편향 문제 (11.4%)	
		키워드	가중치	키워드	가중치	키워드	가중치	키워드	가중치
AI 관련 사건 사고 (4대 토픽)	1	car	0.080	face recognition	0.128	deepfakes	0.030	artificial intelligence	0.133
	2	self driving	0.063	police	0.038	image	0.023	algorithm	0.033
	3	vehicle	0.055	privacy	0.023	Facebook	0.016	bias	0.019
	4	Uber	0.036	government	0.023	face	0.012	chatbot	0.017
	5	autonomous vehicles	0.031	surveillence	0.022	news	0.012	example	0.016

	순위	AI 역기능에 대한 문제제기 (15%)		AI 규제의 필요성 (11.0%)		AI 보안 이슈 (7.5%)	
		키워드	가중치	키워드	가중치	키워드	가중치
AI 관련 한계 및 문제점 (3대 토픽)	1	artificial intelligence	0.091	law	0.028	artificial intellilgence	0.034
	2	robot	0.047	regulation	0.023	security	0.020
	3	human	0.020	rule	0.017	industry	0.015
	4	intelligence	0.014	privacy	0.015	risk	0.014
	5	future	0.013	issue	0.014	threat	0.014

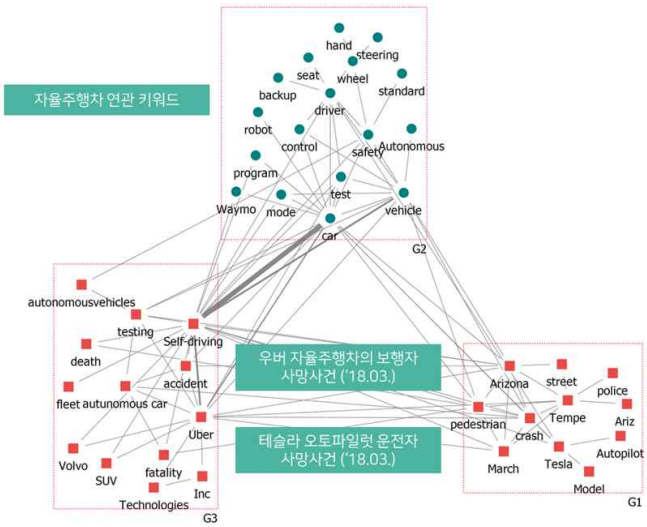
AI 분야의 사건 사고를 주목해보면, △ 자율주행차, △ 안면인식, △ 딥페이크, △ 알고리즘과 챗봇의 네 가지 분야에서 두드러진 현상이 관측되었다. 물론 이 네 가지 토픽 이외의 사건도 있겠으나, 본 연구의 목표는 언론에서 비중있게 다룬 사건을 선별하기 위함에 있다.

이제 각각의 토픽에 해당하는 분석 대상 데이터만으로 키워드 네트워크를 구성한다. 키워든 네트워크는 단락 내용에서 명사만 추출했을 때 인접하는 3개의 명사와 링크를 부여하는 규칙을 적용했다. 이것으로 도출된 키워드 네트워크는 주요 이슈 발굴을 위해 상위 100개 정도 수준의 연결만을 남겨둔 뒤 모두 제거하고 시각화한다. 이를 도식화하면 다음 [그림 2-6]과 같다.

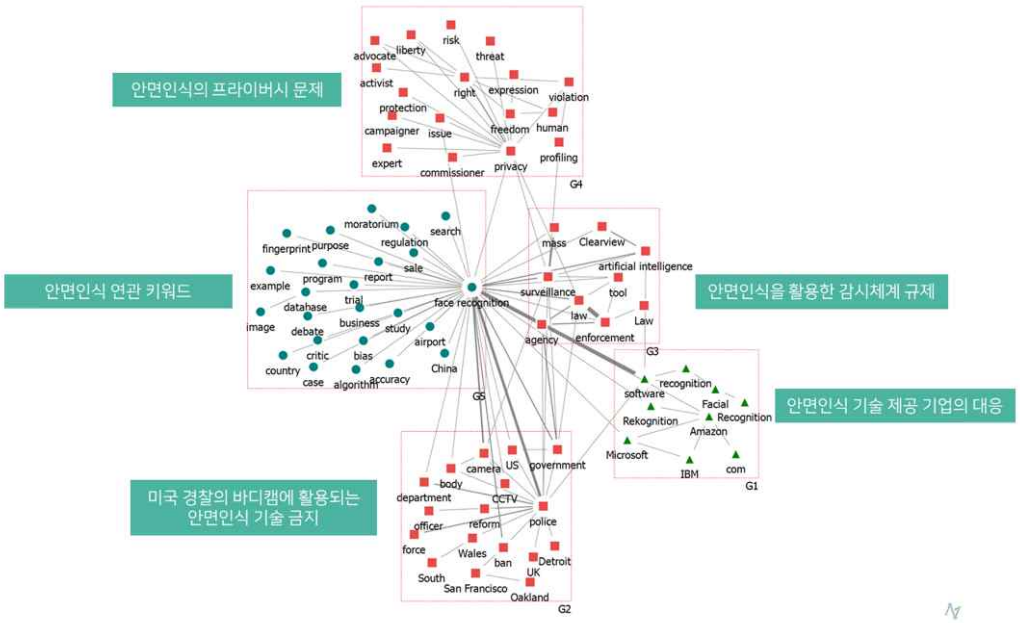


[그림 2-6] 키워드 네트워크 생성 규칙과 시각화

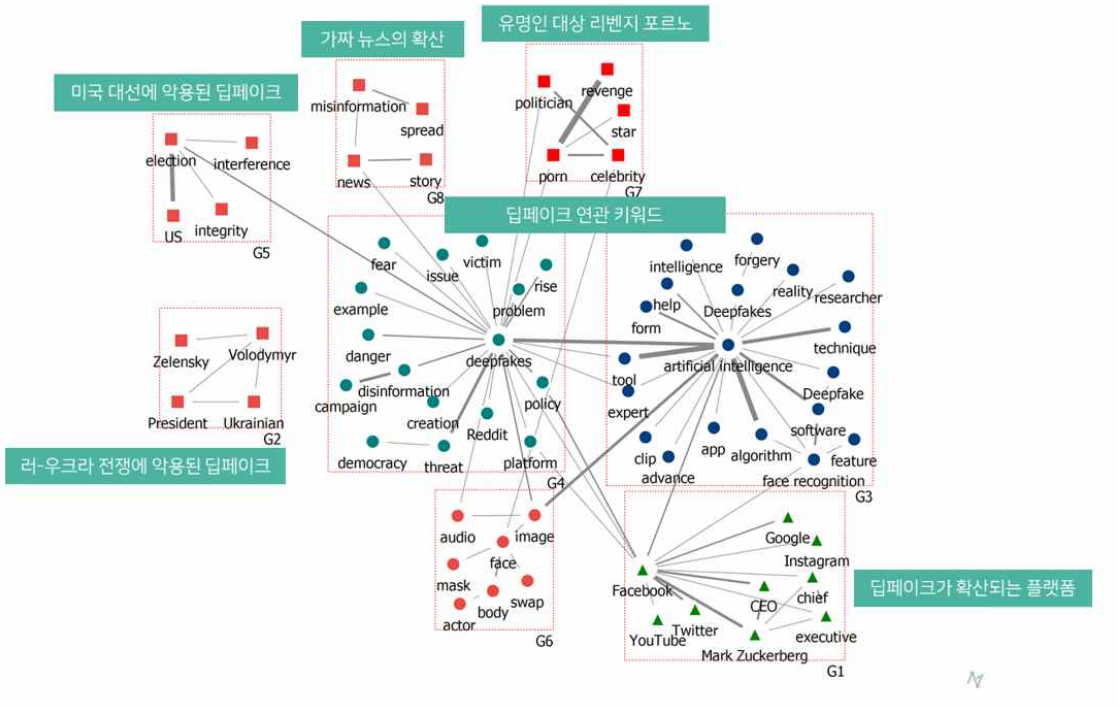
<표 2-8>에서 AI 관련 사건 사고에 해당하는 4대 토픽 각각에 대한 키워드 네트워크 시각화 결과는 다음 그림과 같다.



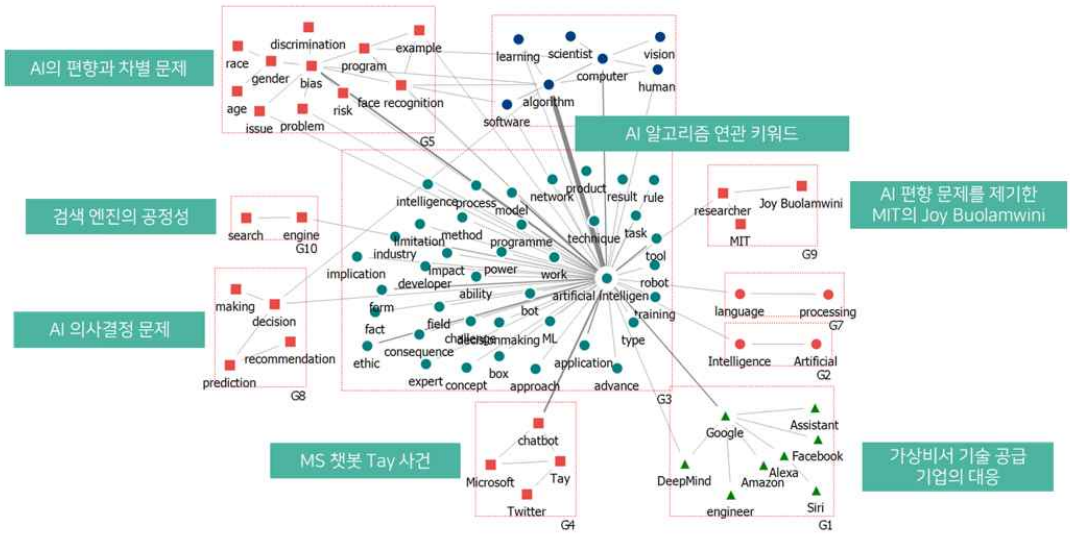
[그림 2-7] AI 토픽 1 - 자율주행 관련 사고



[그림 2-8] AI 토픽 2 - 안면인식 오남용과 감시 체계

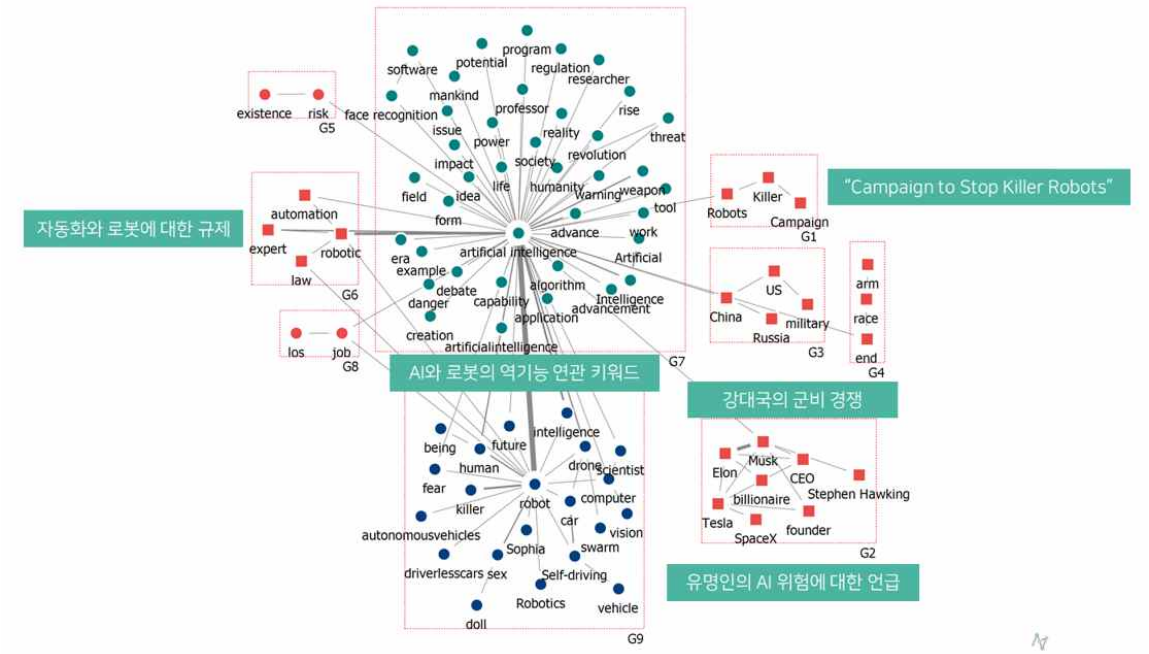


[그림 2-9] AI 토픽 3 - 딥페이크 관련 사건 사고

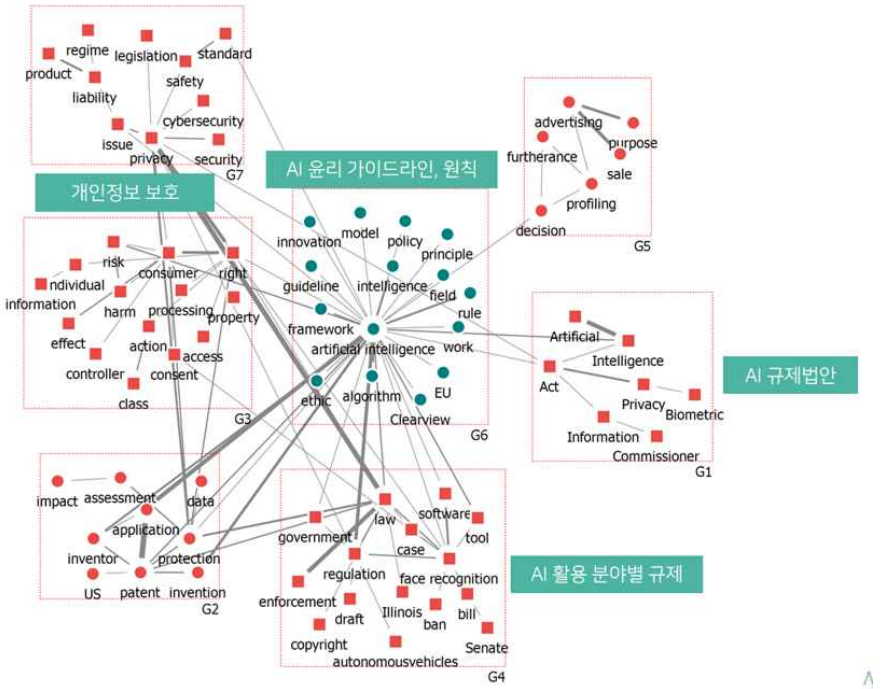


[그림 2-10] AI 토픽 4 - 알고리즘과 챗봇의 편향 문제

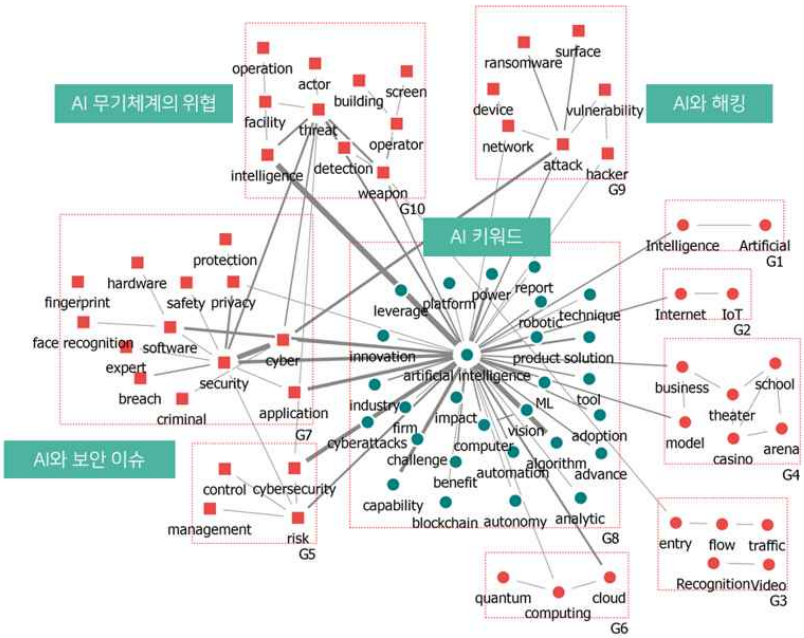
AI 관련 4대 토픽에 대한 키워드 네트워크는 [그림 2-7]부터 [그림 2-10]로 각각 나타나며, 상당히 다양한 사건을 내포하고 있다. AI 관련 사건 사고 토픽에서는 직접적으로 사고를 다루고 있으나, AI의 한계와 문제점 토픽에서는 간접적인 형태로 사건 사고를 다루고 있다. 이 부분 역시 SW 신기술 안전 사고를 도출하는데 활용했다.



[그림 2-11] AI 토픽 5 - AI 역기능에 대한 문제제기



[그림 2-12] AI 토픽 6 - AI 관련 규제의 필요성



[그림 2-13] AI 토픽 7 - AI 보안 이슈

AI 한계와 문제점을 다루는 3가지 토픽은 [그림 2-11]부터 [그림 2-13]과 같다. 본 연구에서는 총 7개의 토픽별로 나타나는 군집이 글로벌 차원에서 주목한 AI 관련 사건 사고로 바라보고 군집의 키워드를 바탕으로 SW 신기술 안전 사고를 도출했다.

다수의 키워드가 군집된 영역을 바탕으로 사건 사고를 추론한 결과, 정성 연구 결과에서 수집한 SW 신기술 안전 사고 목록과 중복되는 것을 제외한 4건의 추가 사고를 도출했다.

다음 <표 2-10>는 메타버스/블록체인 분야에 대한 토픽 모델링 결과이다. 총 6개의 토픽이 도출되었으며 이 중 5개는 블록체인, 1개는 메타버스로 구성된다.

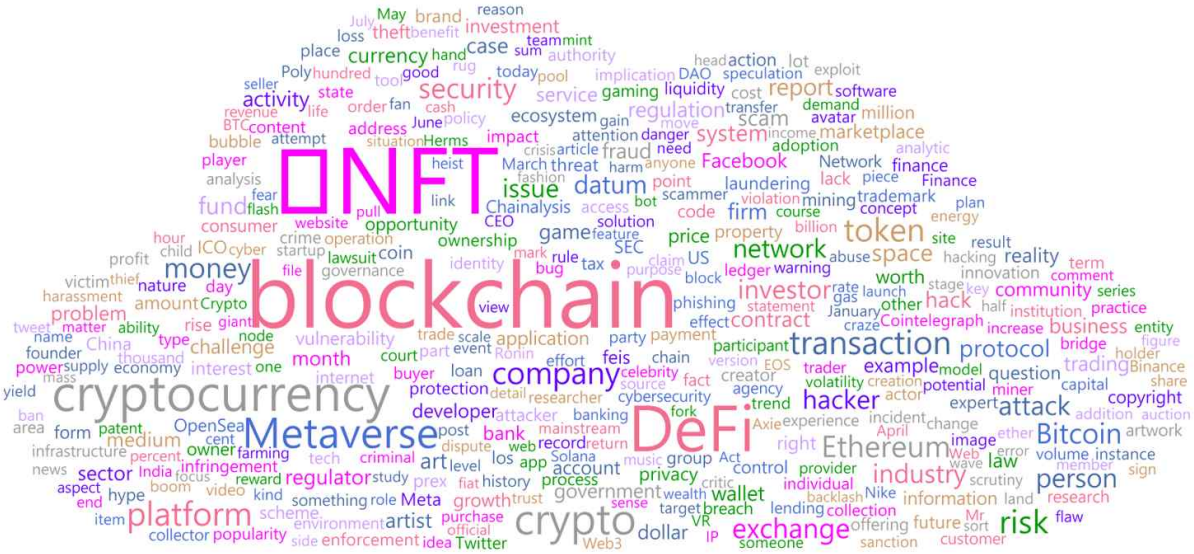
<표 2-10> 메타버스/블록체인 분야 토픽 모델링 결과 - 6대 토픽별 키워드

순위	블록체인 분야					
	블록체인 관련 범죄 (16.7%)		DeFi 논란 (15.5%)		블록체인 규제의 필요성 (16.9%)	
	키워드	가중치	키워드	가중치	키워드	가중치
1	blockchain	0.037	DeFi	0.115	cryptocurrency	0.052
2	hacker	0.027	crypto	0.039	money	0.023
3	cryptocurrency	0.027	token	0.027	blockchain	0.019
4	attack	0.027	protocol	0.024	crypto	0.016
5	exchange	0.023	investor	0.021	regulation	0.016

순위	블록체인 분야				메타버스 분야	
	암호화폐의 위험성 (14.9%)		NFT 관련 이슈 (19.3%)		메타버스 관련 이슈 (16.6%)	
	키워드	가중치	키워드	가중치	키워드	가중치
1	blockchain	0.112	NFT	0.171	metaverse	0.064
2	transaction	0.041	art	0.019	company	0.028
3	Bitcoin	0.028	artisit	0.016	person	0.022
4	network	0.026	marketplace	0.012	reality	0.014
5	Ethereum	0.023	dollar	0.011	game	0.014

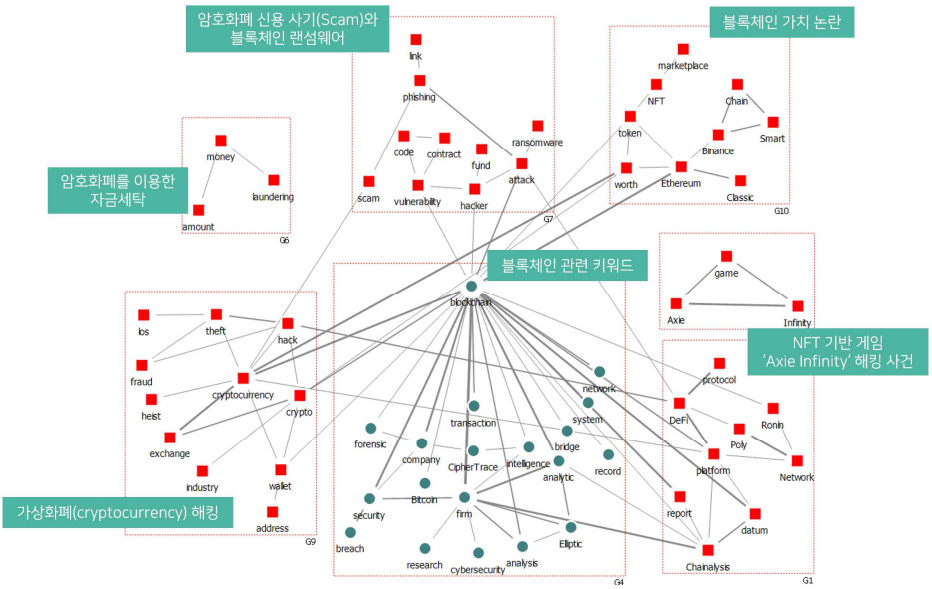
메타버스는 그간 가상현실, 증강현실 등의 키워드로 논의되어오던 영역이 코로나19로 인한 비대면 사회의 확산으로 2020년부터 활발하게 확산세를 이루었다. 반면 블록체인은 대표적인 암호화폐인 비트코인에서부터 출발하여 SW 신기술이지만 메타버스와 비교할 때 상대적으로 긴 역사를 가지고 있다. 따라서 이전 5년간의 언론기사에서는 블록체인의 비중이 상당히 높은 것으로 판단되며, 메타버스는 블록체인과 연관된 문제보다는 기술 자체에서 오는 부작용과 한계에 대한 안전 사고 및 이슈가 주류를 이루었다.

메타버스와 블록체인의 사건 사고, 한계 및 문제점을 언급하는 단락 10,762개에 대해 워드 클라우드를 생성한 결과는 [그림 2-14]와 같다.

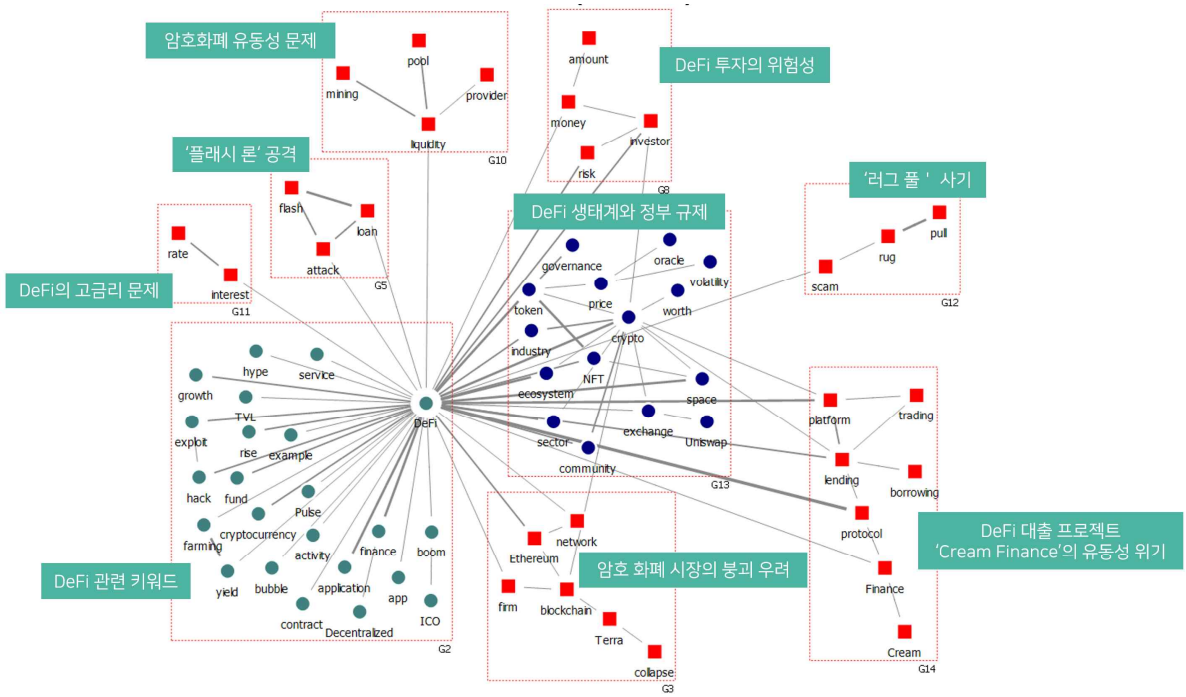


[그림 2-14] 메타버스/블록체인 분야 분석 대상 데이터의 워드 클라우드

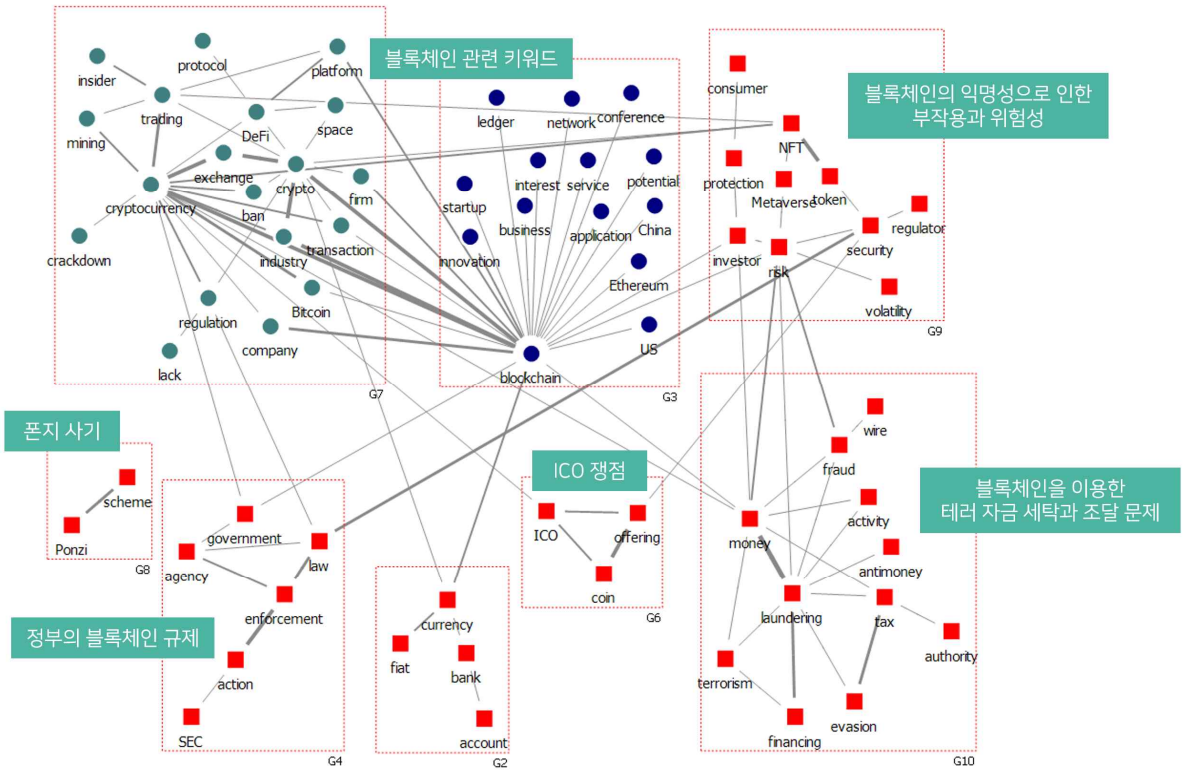
워드 클라우드의 개략적인 경향을 살펴보면 대부분 블록체인과 암호화폐 관련 내용이 대부분을 차지하고 있고, 소수의 영역이 메타버스로 구성되어 있음을 확인할 수 있다. 먼저 블록체인 분야의 5대 토픽에 대한 키워드 네트워크는 다음 [그림 2-15]부터 [그림 2-19]까지로 나타났다.



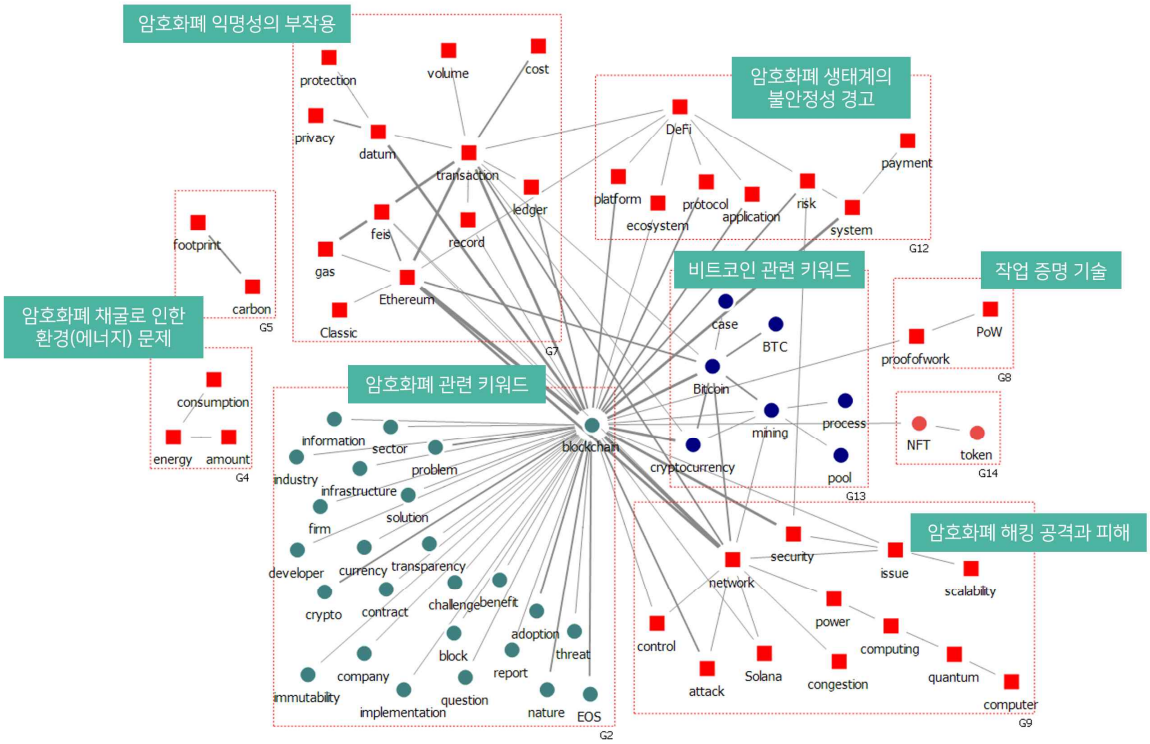
[그림 2-15] 블록체인 토픽 1 – 블록체인 관련 범죄



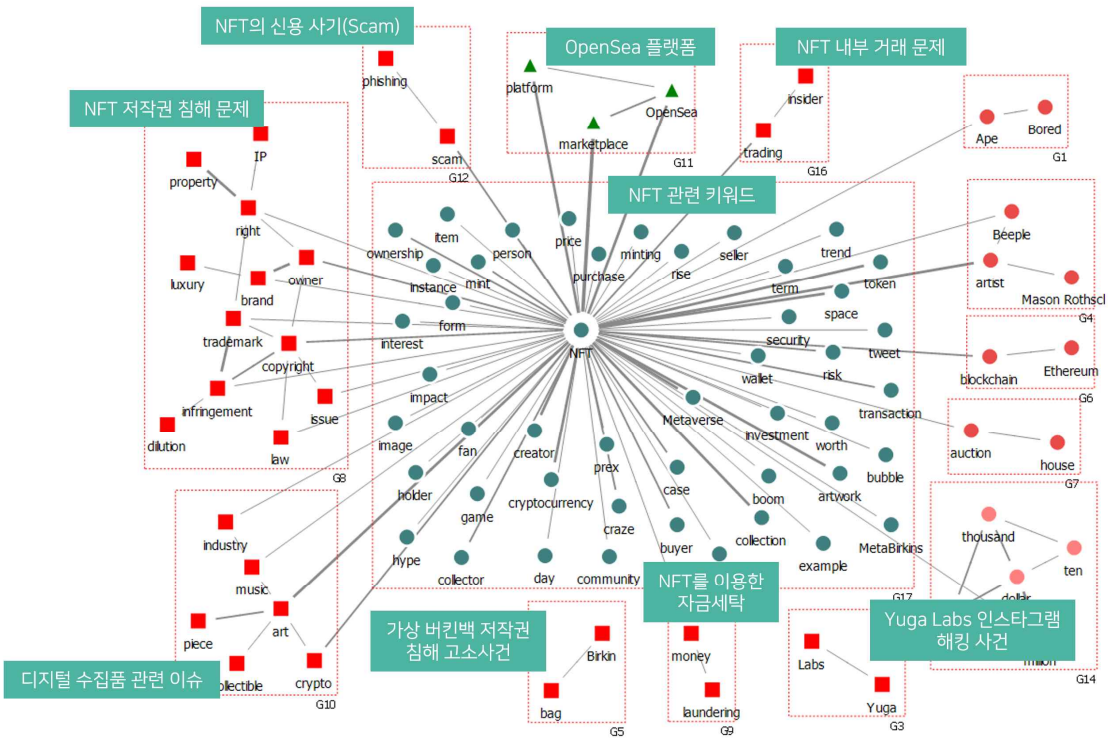
[그림 2-16] 블록체인 토픽 2 - DeFi 논란



[그림 2-17] 블록체인 토픽 3 - 블록체인 규제의 필요성

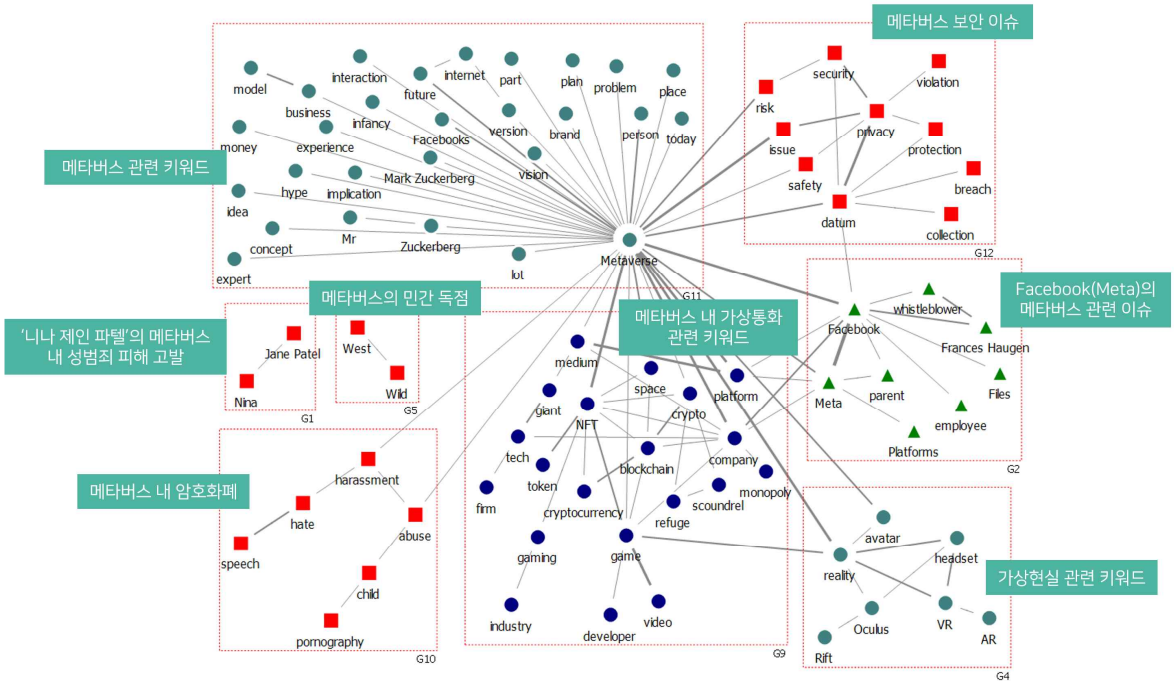


[그림 2-18] 블록체인 토픽 4 - 암호화폐의 위험성



[그림 2-19] 블록체인 토픽 5 - NFT 관련 이슈

메타버스 관련 토픽은 [그림 2-20]과 같다.



[그림 2-20] 메타버스 토픽 - 메타버스 관련 이슈

지금까지 언어모델 재학습을 통해 정제된 5만 여개의 SW 신기술 관련 사건 사고, 한계 및 문제점을 다루는 단락에 대해 토픽 모델링을 적용한 결과 AI 분야에서는 7대 토픽, 블록체인 분야에서는 5대 토픽, 메타버스 분야에서는 1대 토픽이 도출되었다. 이후 각각의 토픽에 해당하는 단락에 대해서 키워드 네트워크를 구축하고 이를 다시 군집하는 분석을 통해 총 13개 토픽에 대한 이슈를 발굴했다.

5. SW 신기술 안전 사고 목록

언론기사 빅데이터 분석의 목적은 정성 연구에서 전문가의 직관에 의존해 추출한 주요 SW 신기술 사건 사고에 대하여 객관성을 확보하기 위한 취지로 진행됐다. 즉, 언론 기사에서 많이 다루고 있는 주요 사건을 선별 및 추가하여 SW 신기술의 안전 사고를 도출하여 최종적인 SW 신기술 안전 사고 보고서를 도출했다.

정성 연구에서 도출된 26건에 더하여 언론기사 빅데이터 분석 결과 추가로 11건의 SW 신기술 사고 보고서를 도출했다. AI 분야에서 4건, 블록체인 5건, 메타버스에서 2건이 추가되었으며 주요 내용의 요약은 다음과 같다.

- AI-15
  - 사고명 : 러시아-우크라이나 전쟁에 악용된 딥페이크
  - 사고 내용 요약 : 2022년 3월 16일 친크렘린 텔레그램 채널 'Operational'은 해커들이 젤렌스키가 비슷한 메시지를 반복하는 딥페이크 영상을 우크라이나 웹사이트에 게시했다고 보도했다. 텔레그램 채널은 동영상 파일을 관리하는데, 메타데이터에 따르면 해당 영상은 3월 16일 텔레그램에 업로드되기 2시간 전에 생성되었다. 특히 티커와 딥페이크 영상에서 공유된 메시지는 우크라이나인들의 항복을 독려하는 내용으로 페이스북과 유사한 크렘린 소유의 소셜미디어 플랫폼인 VK(Vontakte)에서 확산됐다. 러시아의 일부 논평가들은 젤렌스키가 자포자기한 상태에서 영상을 올렸다가 재고 끝에 후퇴했다고 가설을 세웠다.

- AI-16
  - 사고명 : 중국의 신장 이슬람 소수민족 감시체계 강화
  - 사고 내용 요약 : 2016년 말부터 중국 정부는 신장에 있는 1300만 명의 위구르족과 투르크계 무슬림들에게 대량의 자의적 구금, 강제적인 정치적 세뇌, 이동제한, 종교적 억압을 가했다. 실제 정확하게 드러난 사실로, 중국 정부의 억압하에 최대 100만 명의 사람들이 "정치 교육" 수용소에 수용되어 있다. 중국 정부의 강압적이고 폭력적인 테러에 대항하는 시민들의 강경 시위는 신장을 사회 통제를 위해 혁신적인 기술인 AI을 사용하는 중국의 주요 중심지 중 하나로 만들었다. 2015년 공안부는 2020년까지 전국에 완전한 영상 감시 네트워크를 구축하겠다고 발표했으며 2016년 이후 5천명 이상의 범죄자를 체포했다고 밝혔다.

- AI-17
  - 사고명 : 아마존 AI 채용 도구의 성별 편향
  - 사고 내용 요약 : 2018년 10월 11일, 글로벌 기업 아마존에서 AI 채용 프로그램 논란이 발생했다. 아마존에서 자체 AI 채용 프로그램을 개발했는데, 실제 채용에 적용하기 전 최종 시뮬레이션 과정에서 남성 지원자가 여성 지원자보다 평균적으로 더욱 높은 점수를 받는 편향 현상이 발견되었다. 원인은 아마존의 직원 구성으로 밝혀졌다. 아마존은 기본적으로 IT 기업으로써 개발직군이 70%를 차지했으며 개발직군의 남자 직원이 여성 직원보다 압도적으로 많은 인

원이였다. 그에 따라 남성 직원 중에 고성과자가 많았다. AI 채용 프로그램은 아마존 내부 직원들의 데이터를 학습하고 기준으로 삼았기 때문에 해당 모수를 곧 채용의 기준이 되었다. 2021년 2월 11일, 결국 아마존은 데이터를 수정하거나 알고리즘을 수정하는 것으로는 한계가 있다고 판단하여 AI 채용 프로그램을 폐기하기로 결정했다.

- AI-18
  - 사고명 : 범죄예측 프로그램 COMPAS의 인종 편향
  - 사고 내용 요약 : 비영리단체 프로퍼블리카는 COMPAS를 분석한 결과, 흑인 피고인이 백인 피고인보다 재범 위험성이 높은 것으로 잘못 판단될 가능성이 훨씬 높은 반면, 백인 피고인은 흑인 피고인보다 낮을 가능성이 더 높은 것으로 판단하고 있다는 사실을 발견했다. 플로리다 브로워드 카운티에 있는 1만 명 이상의 형사 피고인들을 조사했고 그들의 예상 재범률을 2년 동안 실제로 발생한 비율과 비교한 결과, COMPAS 도구에 의해 예측된 재범 위험 범주와 점수가 매겨진 후 2년 동안 피고인의 실제 재범률을 비교한 결과 점수는 가해자의 재범률을 61%로 정확하게 예측했지만 폭력적 재범률을 20%로 예측한 경우에만 정확한 것으로 나타났다. 누가 재범할지 예측할 때, 알고리즘은 흑인과 백인 피고인의 재범률을 거의 같은 비율로 정확하게 예측했지만(백인 피고인 59%, 흑인 피고인 63%) 매우 다른 방식으로 실수를 했다. 분석 결과, 과거 범죄, 재범, 연령, 성별 등을 통제할 때에도 흑인 피고인이 백인 피고인보다 위험 점수가 더 높게 배정될 가능성이 45% 더 높았기 때문이다.

- Metaverse-08
  - 사고명 : 메타버스 속 집단 성범죄 피해
  - 사고 내용 요약 : 영국 여성 ‘니나 제인 파텔(Nina Jane Patel)’이 메타버스 속에서 집단 성범죄를 당했다고 밝혔다. 테크 기업 카부니 벤처스(Kabuni Ventures)의 메타버스 연구 부사장 니나 제인 파텔이라고 이름을 밝히고 공론화를 한 피해 여성은 “메타(Meta)의 가상 게임 호라이즌 월드(Horizon Worlds)에서 가입한지 60초 만에 남성 아바타와 남성 목소리를 한 다른 사용자 3~4명이 범한 집단 성범죄 피해를 당했다. 그들은 사실상 갱단과 같았다.”라고 작성했다. 이어서 “가해자는 나의 아바타 사진을 촬영하고 ‘성적 행위를 싫어

하는 척 하지 마라’ 등과 같은 메시지를 보냈다” 라고 덧붙였다. 그녀는 그것을 “너무 빨리 일어난 끔찍한 경험“이라고 묘사하여 “안전 장치” 를 사용할 생각을 할 시간조차 없었다. 파텔은 현재 미국과 캐나다 18세 이상 사용자 누구나 호라이즌 월드에 접속할 수 있다는 점과 최근 들어 메타버스 성범죄 피해 사례가 여러 차례 보고된 점을 지적하며, 문제의 심각성을 강조했다.

- Metaverse-09
  - 사고명 : 트렌드마이크로, 메타버스 악용의 공간 ‘다크버스’ 경고
  - 사고 내용 요약 : 다크버스는 모두가 이용하는 사이버공간의 허점을 파고들어 본인들만의 이익을 취하는 악성 공간으써 활용되고 있다. 다크버스는 메타버스 내부에 존재한다는 점을 제외하면 다크웹과 비슷하다. 다른 방면에서 다크버스를 봤을 때 평범한 사용자는 사이버상의 존재와 물리적 존재에게 동시에 영향을 받는 사이버-물리 때문에 다크웹보다 더 위험하다. 사이버-물리란 가상세계와 물리적 실체가 연동되는 것으로, 가상의 영역에 속하는 컴퓨팅, 통신, 제어를 실제 물리적 세계와의 통합을 의미한다. 다크웹 범죄 포럼의 온라인 공개 토론의 장에서는 다양한 위협들에 대해 끊임없이 논의되고 있다.

- Blockchain-06
  - 사고명 : NFT 기반 게임 ‘엑시 인피니티’ 해킹
  - 사고 내용 요약 : 2022년 3월 29일 블록체인 기반 게임 엑시 인피니티가 해킹당하는 사건이 발생했다. 해커들이 엑시 인피니티에서 사용되는 암호화폐 이더리움과 USDC를 6억 2,500만 달러어치를 빼돌린 사건이다. 해커들은 엑시 인피니티를 운영하는 스카이마비스가 개발한 사이드체인 로닌 네트워크를 침입해 해킹을 진행했다. 이어서 2022년 4월 14일 미국 재무부는 엑시 인피니티 해킹 사건에 북한 해킹 조직 ‘라자루스’가 연계되어 있다고 발표했다. 그에 대한 조치로 라자루스와 관련되어 있다고 추정되는 암호화폐 이더리움 계좌를 차단하고 중지시키는 조치를 취했다고 밝혔다. 라자루스는 북한 인민군 정찰총국과 연계되었다고 알려진 해커 조직으로써 2014년에 김정은 북한 국무위원장을 조롱한 영화를 제작한 미국 영화사와 소니픽처스를 해킹하는 사건을 일으켰었다.

- Blockchain-07

- 사고명 : 가상 버킨백 만든 예술가, 에르메스에 피소
- 사고 내용 요약 : 프랑스 명품 브랜드 에르메스에서 '가상 버킨백'을 만들어서 판매한 미국 예술가를 상대로 소송을 낸 사건이 발생했다. 2022년 1월 22일 에르메스는 예술가 메이슨 로스차일드에 대해 자사의 상표로 이익을 얻으려 한다는 이유로 미국 뉴욕법원에 고소장을 제출했다. 로스차일드는 지난달 NFT 거래소 오픈시(OpenSea)에 에르메스에서 만들었던 버킨백과 유사한 디지털 작품을 출품했다. 해당 작품은 '메타 버킨스(Meta Birkins)'로 실제 에르메스에서 만든 버킨백의 그림파일에 다양한 소재와 색을 입히는 형식으로 작품을 구성했다. 실제 에르메스의 버킨백과 유사한 해당 작품에 많은 사람이 관심을 가졌고 10억원 이상의 가격이 측정되어 판매될 만큼 인기를 끌었다.

• Blockchain-08
- 사고명 : 크림 파이낸스, 플래시론 공격으로 1,500억원 손실
- 사고 내용 요약 : 2021년 10월 27일, DeFi 대출 프로젝트 크림 파이낸스(Cream Finanace)에서 플래시 론 공격으로 인해 약 1억 3000만달러의 가상화폐 탈취 피해가 발생했다. 공격자는 주로 크림파이낸스의 유동성공급(LP) 토큰인 크림(CREAM)과 이더리움(ERC-20)기반 토큰들을 탈취했으며, 이 과정에서 68종 토큰을 대상으로 9이더리움(ETH) 이상의 수수료를 사용한 것으로 알려졌다. 해당 사건으로 인해 이더리움 기반으로한 크림 파이낸스의 유동성풀이 동결되었으며 복구는 불가능 하다고 알려졌다. 크림 파이낸스가 외부 공격으로 인해 가상화폐 손해를 보게된 세 번째 사건이다. 크림파이낸스는 2021년 8월 같은 플래시 론 공격으로 1880만 달러를, 2021년 2월 3750만 달러 상당을 피해본 사례가 있다. DeFi 자금손실을 다루는 REKT가 발표한 자료 기준 이번 사건은 역대 디파이 자금탈취 공격 중 세 번째로 큰 규모다.

• Blockchain-09
- 사고명 : 중국, 모든 가상화폐 거래 불법화 발표
- 사고 내용 요약 : 2021년 9월 24일 중국의 중앙은행인 중국인민은행이 모든 종류의 가상화폐 거래를 불법으로 규정하고 엄격한 단속을 시작하기로 발표했다. 중국인민은행은 "가상화폐는 법정 화폐와 동등한 법적 지위를 보유하지 않는다. 가상화폐 관련 업무 활동은 불법적인 금융 활동에 속한다"라고 밝혔다.

이때 가상화폐 관련 업무에 포함되는 범위는 가상화폐 거래에 정보를 중개하고 정해진 돈을 받는 모든 서비스와 파생상품 거래, 불법 모금도 모두 포함된다. 2020년까지 세계 가상화폐 채굴의 65%를 차지한 중국에서 이러한 파격적인 발표를 하게 되었다.

- Blockchain-10
  - 사고명 : 암호화폐 채굴 장려하던 카자흐스탄, 전력난 심각
  - 사고 내용 요약 : 2021년 11월 29일, 정부 차원에서 암호화폐 채굴을 장려하던 카자흐스탄이 심각한 전력난을 겪고 있는 것으로 알려졌다. 카자흐스탄 전력공급회사는 최근 정부에 등록된 채굴업자 50명에게만 전력을 공급하기로 결정했다고 발표했다. 이는 2021년 10월 카자흐스탄 내 석탄 화력 발전소 3곳을 폐쇄된 이후 전력 수급이 원활치 않기 때문이다. 카자흐스탄 에너지부는 2021년 말까지 전력 수요가 작년 대비 1~2% 증가 수준이 아닌 8%까지 급증했다고 추정했다. 카자흐스탄에서는 지난 10월 이후 현재까지 6개 지역에서 정전이 발생했다. 카자흐스탄 정부와 전문가들은 카자흐스탄 지역에서 정부에 등록하지 않고 집이나 공장에서 불법적으로 채굴하는 암호화폐 채굴자들이 급증하면서 정전 사태가 발생하는 것이라고 지적하고 있다. 또, 최근 중국이 암호화폐와의 전쟁을 벌이자 채굴자들이 암호화폐 채굴을 장려하는 카자흐스탄으로 대거 옮긴 까닭도 있다.

제4절 SW 신기술 안전 사고 조사 결과

SW 신기술로 인해 발생한 안전 사고 조사는 SW 안전 전문가가 직접 선별한 26건의 사고와 언론기사 빅데이터 분석을 통해 추가로 도출한 11건의 사고를 더하여 총 37건의 SW 신기술 안전 사고 조사 보고서를 확보했다. 특히, 언론기사 빅데이터 분석 결과에서는 총 25건의 사고를 선별했는데, 이중 정성 연구에서 선별된 결과와 14건이 중복되어 최종적으로는 11건의 사고를 더했다. 총 37건의 SW 신기술 안전 사고 목록은 <표 2-11>과 같다.

<표 2-11> SW 신기술 안전 사고 목록 (총 37건)

연 번	사고 제목
AI-01	신약 개발 AI의 악용 가능성 제기
AI-02	AI 딥페이크 기술을 이용한 여성 나체사진 유포
AI-03	자율주행차량에 의한 첫 보행자 사망사고
AI-04	숨진 가족 목소리 되살리는 아마존 AI '알렉사'의 신기능 논란
AI-05	AI 챗봇 '이루다' 의 혐오 발언 및 개인 정보 유출로 인한 서비스 중단
AI-06	안면인식 AI 프로그램 오류로 무고한 흑인 남성 체포
AI-07	학습데이터 편향성으로 인한 AI 학력평가 알고리즘의 오류
AI-08	AI 의사 '왓슨(Watson)', 국내 실정과 맞지 않는 의료 AI의 한계
AI-09	음성 딥페이크 기술을 이용한 보이스피싱
AI-10	日 유명 호텔, 잦은 고장과 오류로 로봇 종업원 해고
AI-11	로봇의 오작동으로 인해 체스 로봇과 대국하던 소년이 당한 신체 부상
AI-12	드론 오폭으로 인한 민간인 사망 사고
AI-13	유튜브 추천 알고리즘으로 인한 편향성 강화
AI-14	배송 드론 추락사고
AI-15	러시아-우크라이나 전쟁에 악용된 딥페이크
AI-16	중국의 신장 이슬람 소수민족 감시체계 강화
AI-17	아마존 AI 채용 도구의 성별 편향

AI-18	범죄예측 프로그램 COMPAS의 인종 편향
Metaverse-01	미성년자 대상 디지털 성범죄
Metaverse-02	현실과 비슷해지는 디지털 성범죄
Metaverse-03	메타버스 공간 내 지적 재산권 침해
Metaverse-04	가상부동산 서비스 중단으로 인한 유료 구매자의 금전적 손실
Metaverse-05	가상세계 과몰입으로 겪는 신체 부상
Metaverse-06	AR 게임 포켓몬 고로 인해 발생한 교통사고 및 안전사고
Metaverse-07	메타버스의 플랫폼화를 가로막는 주요 난제 중 하나인 VR 울렁증
Metaverse-08	메타버스 속 집단 성범죄 피해
Metaverse-09	트렌드마이크로, 메타버스 악용의 공간 '다크버스' 경고
Blockchain-01	크로스 체인 브릿지에서 발생한 대규모 해킹
Blockchain-02	알고리즘 기반 스테이블 코인의 폭락으로 인한 투자자 손실
Blockchain-03	NFT 거래소 해킹
Blockchain-04	블록체인 기반 NFT 신종 사기
Blockchain-05	솔라나 네트워크 장애
Blockchain-06	NFT 기반 게임 '엑시 인피니티' 해킹
Blockchain-07	가상 버킨백 만든 예술가, 에르메스에 피소
Blockchain-08	크림 파이낸스, 플래시론 공격으로 1,500억원 손실
Blockchain-09	중국, 모든 가상화폐 거래 불법화 발표
Blockchain-10	암호화폐 채굴 장려하던 카자흐스탄, 전력난 심각

언론기사 빅데이터 분석 결과에서는 직접적인 사건 사고와 더불어 정부의 대응이나 기관의 경고 메시지 등이 주요 이슈로 부각되었기 때문에, 이를 사고 보고서에 추가한 점이 차별화된 시각이라 볼 수 있다. 이러한 내용을 추가한 이유는 후술하게 될 SW 신기술 위험원 유형화에서도 유의미한 결론을 얻을 수 있기 때문이다.

제3장 SW 신기술 위험원 유형화 모델

제1절 개 요

1. SW 신기술 위험원 유형화 고려사항

SW 신기술에 대한 사고 유발 위험원을 유형화하고자 하는 궁극적인 목표는 SW신기술에서의 잠재 안전 사고에 대한 예방에 있다. 예방이란, 잠재 위험을 예측하고 이를 피하는 노력을 의미한다. 이를 위해서는 어떠한 잠재 위험원들이 있는지 예측할 수 있어야 한다.

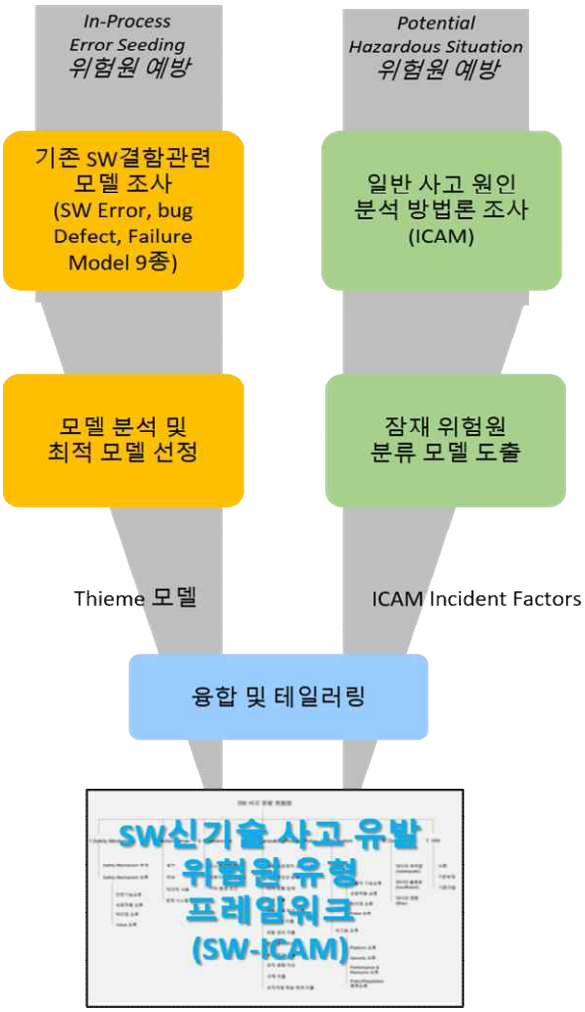
SW의 위험원이란 SW 공학적으로 크게 두 가지로 분류될 수 있다. 첫 번째는 기능 혹은 비기능 요구사항을 충족하도록 시스템을 개발할 때 전파되는 SW의 오류로 인하여 발생하는 위험이다. 후술할 SW 결함 분류에 관련된 연구들은 이러한 위험원에 대한 SW 공학적인 유형화 혹은 분류 체계에 대한 것이다. 통상적으로 이러한 연구는 범도메인(cross-domain)특성과 개발 기술(개발 패러다임)에 의존적인 특성을 가지고 있다.

두 번째는 단순히 SW 개발 과정상에서 전파되는 오류가 아니라, 개발하는 SW가 작동 하는 과정상 잠재 위험이 있는 조건과 상황들을 고려하지 못해서 발생하는 사고의 위험들이다. 다시 말하자면 SW 개발 시 안전 요구사항을 식별할 때 식별 자체를 못하여 SW 개발에 반영되지 못한 상황을 의미한다. 두 번째 위험원에 대한 식별을 위해서는 일반적으로 특정 도메인에 대한 축적된 지식이 필요하다. 자동차의 특정 부품, 항공기의 특정 부품 등 해당 부품과 작동하는 상황에 대한 축적된 지식이 있어야만 잠재 위험원을 예측하고 도출할 수 있는 것이 일반적인 논리이다.

기능 안전 분야에서 SW 안전 공학이 자리 잡은 산업 분야의 경우 상기 두가지의 위험원 분석이 모두 적용되고 있다. SW이기 때문에 첫 번째 분류의 위험원에 대한 분석은 당연히 적용되고, 특정 도메인 전문지식을 통한 두 번째 분류의 위험원 분석도 적용하고 있다. 도메인별로 기능 안전 표준이 적용 분야들을 보게 되면 자동차, 항공, 철도와 같이 그 역사가 오래되어 그동안 축적된 위험원에 대한 전문지식들이 존재하고 있다.

2. SW 신기술 위험원 유형화 전략

SW신기술의 경우, 위에서 언급한 첫 번째 분류의 위험원에 대한 분석은 당연히 적용되어야 한다. 다만 두 번째 분류의 위험원의 경우 오랜 시간을 통하여 축적된 위험원 지식이 있어야 하는데, SW 신기술이다 보니 그러한 데이터나 지식이 부재한 상황이다. 단순한 논리로만 본다면, 절대적 시간을 갖고 SW 신기술에서 발생하는 사고들을 지속적으로 분석 및 축적하여야 한다. 하지만 그러기에는 SW 신기술이 발전하고 확산되는 속도가 너무나 빠르다. 위험에 대한 지식 축적 속도보다 신기술이 새로운 분야로 확산되는 속도나 기술의 진화가 훨씬 빠르다는 것이다. 따라서 전통적인 접근 자체가 현실적으로 어렵다고 볼 수 있다. 그러므로 안전에 대한 요구사항을 빠짐없이 잘 식별해야 하는 두 번째 분류의 위험원에 대한 새로운 접근이 필요하다.



[그림 3-1] SW-ICAM의 논리적 개발 흐름

본 연구에서는 이를 위해서 안전 요구사항에 대한 메타 레벨(meta level)의 유형화를 모색하였다. 특정 도메인의 위험원 유형을 찾기보다는, 일반적인 사고의 위험원 유형을 찾아보도록 하는 것이다. 우리가 일반적인 사고의 위험원 유형을 SW 관점에서 사용할 수 있다면 SW 신기술과 같이 기술의 진화 속도가 빠르고, 또한 실제 적용한 역사가 짧은 분야에 매우 유용하다고 볼 수 있다. 이를 위하여 일반화된 사고 원인 조사를 하는 방법론, 혹은 이러한 방법론에서 사용하는 사고 원인 유형 모델을 조사했으며, 대표적인 사고 원인 조사 방법론으로서 영국에서 개발된 ICAM(Incident Cause Analysis Method)을 참조하고, ICAM에서 제시하는 사고 원인 유형 분류 체계를 활용하도록 하였다.

이를 통하여 본 연구에서는 개발 과정상의 오류 전파 특성을 가진 결함분류체계의 대표적인 모델과 ICAM에서 제시하는 잠재 사고 원인(위험원)인 모델을 합성하고 재단하여 새로운 SW-ICAM이라는 SW신기술 사고 유발 유형 모델을 개발하였다. 이것은 전통적인 사고 원인 분석 방법론이라기 보다는, 잠재 사고 위험원 추론을 위한 위험원 유형 프레임워크이다. [그림 3-1]은 지금까지 설명한 SW-ICAM의 논리적 개발 흐름을 나타낸 것이다.

제2절 선행연구 분석

1. SW 결함 분류

SW 안전사고는 대부분 SW 결함으로부터 기인한다. SW 결함(software fault)은 일반적으로 명세(specification), 프로그래밍, 설치(installation)와 같은, SW 개발 단계 중에 SW에 도입된 SW에 설계 결함을 말한다(Jean-Claude et al., 1995; Littlewood and Strigini, 2000).

"Research Issues in Software Fault Categorization" (Ploski et al. 2007)연구에서는 서로 다른 SW 개발 프로젝트의 SW 결함 사이에는 어떤 유사점이 있고, 어떤 분류(categorization)가 적용되었는지 기존 연구들을 분석하였다. 버그 분류(bug taxonomies), SW 결함 분류(software defect classification) 및 근본 원인 결함 분석(root causal defect analysis) 등을 포함한다.

이전 연구들에서 결함을 분류하려는 목적은, 프로그래밍 실수의 근본 원인을(root cause) 이해하기 위해서, 결함 감내(fault tolerance) 측도(measures)를 개발하기 위해서, 테스팅을 용이하게 하기 위해서, SW 개발 프로세스를 모니터링하기 위해서, 자동화를 통해 개선하기 위해서 등 다양했다.

<표 3-1> 기존 SW 결함 분류 스키마 개요 (Ploski et al., 2007)

주저자	Knuth (Knuth, 1989)	Beizer (Beizer, 1990; Beizer and Vinter, 2001)	Gray (Gray, 1986)	Chillarege (Chillarege et al., 1992)	Eisenstadt (Eisenstadt, 1997)	DeMillo (DeMillo and Mathur, 1995)
1 목적	교육	테스팅	결함 감내	프로세스 개선	디버깅	테스팅
2a 프로젝트 개수	1	multiple	2000	>50	multiple	1
2b 프로젝트 사이즈	1-8 devs/10 yrs	large	large	large	-	1-8 devs/10 yrs
2c 결함개수	867	982	132	-	51	415
2d 적용 도메인	타입세팅	국방 항공 통신	트랜잭션 처리	시스템 SW	지정되지 않음	타입세팅

3 정보 출처	개발자	개발자 테스터	고객	개발자 테스터	개발자	구문 분석
4 샘플 결함 속성들	결함 도입 시점 소스에서의 위치 발현 실수 타입 회피가능성	요구되는 수정 활성화 영향 위치 결함 도입 시점	일시적임	요구되는 수정 활성화 트리거 탐지 시점	소스에서의 위치 실수 타입 활성화 트리거	소스에서의 위치 요구되는 수정
5 카테고리의 의도된 배타성?	yes	yes	yes	no	yes	yes
6 모호성(ambiguity) 위험	높음	높음	낮음	중간	높음	낮음
7a 데이터 가용여부?	Knuth	Beizer	Gray	Chillarege	Eisenstadt	DeMillo
7b 결함 카테고리별 빈도 데이터 가용여부?	yes	yes	yes	no	yes	yes
8 추천 프로세스?	없음	없음	있음	있음	없음	있음
9 말단 카테고리 개수	9	>100	2	8+18	10	>20
10 카테고리 구성	계층없음 (flat)	계층적	계층없음	계층없음	계층없음	계층적

- Knuth 연구 (Knuth, 1989)
  - TEX 시스템에서 발생한 오류들(errors)을 아래와 같은 결함 클래스로 분류하였다.
    - ✓ A 잘못된 알고리즘
    - ✓ B 실수
    - ✓ D 데이터 구조 와해
    - ✓ F 잊어버린 기능
    - ✓ L 언어 책임

  - ✓ M 모듈 간 불일치
  - ✓ S 놀라운 시나리오
  - ✓ T 사소한 오타
- Knuth 연구에서는 다음과 같은 결함 속성(attributes)들을 암묵적으로 고려하였다.
  - ✓ 결함이 도입된 시간 (예, A - 알고리즘 명세 중, B - 코딩 중)
  - ✓ 결함 위치 (예, D - 데이터 구조: 결함이 애플리케이션 모듈에도 할당됨),
  - ✓ 소스 코드의 표현 또는 그것의 부족 (커미션 및 누락 오류, 예, F - 잊어버린 기능)
  - ✓ 프로그래머가 잘못 해석한 정보의 종류 (예, M - 모듈 인터페이스, L - 언어 규칙)
  - ✓ 결함이 도입되는 시점에 결함을 회피하는 프로그래머의 능력 (예, B - 쉽게 회피할 수 있는 실수, S - 예측하기 어려운 미래 시나리오).
- Knuth 연구에서 결함 분류는 프로그래밍 실수의 근본 원인을 이해하고자 하는 의도로 수행되었으며, 전반적으로 분류가 임의적이고 모호성 많다.

• **Beizer의 버그 분류(bug taxonomy) (Beizer, 1990; Beizer and Vinter, 2001)**
- Beizer의 연구에서는 8가지 상위 수준 결함 카테고리를 제안하였다.
  - ✓ 기능적 버그: 요구사항 및 피처(features) (요구사항 명세 중 오류)
  - ✓ 구현된 기능 (요구사항 해석의 오류)
  - ✓ 구조적 버그 (제어 흐름 및 표현 조작에 관한 구현 중 실수)
  - ✓ 데이터 버그 (데이터의 정의, 구조, 또는 사용에서의 버그)
  - ✓ 구현 (인쇄상 버그, 표준 및 규칙 위반, 문서에서의 오류)
  - ✓ 통합 (컴포넌트 간의 인터페이스와 관련된 버그)
  - ✓ 시스템 및 SW 아키텍처
  - ✓ 테스트 정의 및 실행
- Beizer의 연구에서는 상위 수준 결함 카테고리 아래 여러 하위 카테고리를 제안하여, 결함 카테고리가 계층구조를 가지고 있다.
- 결함 카테고리 간에 완전 독립적이 아니며, 정의에 모호성 많다.

• **Gray의 생산 SW 대상 SW 결함 분류 연구 (Gray, 1986)**
- 실행 시간(run time)에 결함 감내를 잘하기 위한 목적으로 SW 두 가지 부류로 분류하였다.

  - ✓ 일시적 결함(transient faults): 일관된 초기 상태와 재시도를 복원하여 마스킹할 수 있는 결함으로(원본 논문에서 Heisenbugs 라고 명명), 반복적인 실행으로 간단하게 재현할 수 없다.
  - ✓ 비일시적(non-transient faults): 재시도 함으로써 활성화를 마스킹할 수 없는 결함으로(원본 논문에서 Bohrbugs라고 명명) 쉽게 재현 가능하다.
- 생산 SW에만 집중하였고, 실행 시 활성화가 되는지 여부에 따라 결함을 분류하였다.
- 결함 감내 메커니즘을 상세화하는 분야에서는 유용하겠지만, 보다 큰 목적을 고려하면 충분하지 않은 분류이다.

• Chillarege 외의 Orthogonal Defect Classification 연구 (Chillarege et al., 1992)

- Chillarege 외의 연구[Chillarege et al. 1992]에서는 SW 결함의 수를 줄이기 위한 목적으로, 다양한 SW 개발 단계에서 개발자에게 피드백을 주려는 의도에서 Orthogonal Defect Classification(ODC) 개념을 소개하였다.
- SW 수정(fix)에 대한 정보는 결함 유형(defect type)이라는 속성(attribute)으로 인코딩되고, 결함 트리거(defect trigger) 속성은 SW 결함이 활성화되는 상황을 설명하는 데 사용된다.
- 다음과 같은 결함 타입들을 정의하였다.
  - ✓ 기능 (누락 되거나 잘못된 기능, 정형적인 설계 변경이 필요할 수 있음)
  - ✓ 인터페이스 (사용자, 모듈 또는 장치 드라이버 간의 통신 오류를 다룸)
  - ✓ 체킹 (소스 코드에서 데이터와 값의 유효성 검증이 잘못되거나 누락됨)
  - ✓ 할당 (잘못된 초기화와 같은 소스 코드에서의 결함을 다룸)
  - ✓ 타이밍/직렬화 (공유 리소스 및 실시간 리소스 관리를 개선하면 수정되는 오류)
  - ✓ 빌드/패키지/병합 (라이브러리 시스템에서의 실수, 변경 관리 및 버전 제어로 인한 문제를 다룸)
  - ✓ 문서화 (출판물과 유지 관리 노트를 다룸)
  - ✓ 알고리즘 (설계 변경을 요청할 필요 없이 재구현하여 해결할 수 있는 효율성 또는 정확성 문제를 다룸)
- Chillarege 외의 연구에서의 분류는 SW 결함의 본질을 다루기보다는 실무에서 사용하기 위한 의도가 강하다.
- ODC가 SW 결함 만에 국한되지는 않는다는(예, 기능적 결함) 특징이 있고,

ODC는 SW 개발 ‘프로세스’를 특징짓고 개선하는데 유용하다. SW 결함 정의를 ‘산출물에서의 결함’에서 ‘프로세스에서의 결함’으로 확장한 의의가 있다.

- Chillarege 외 연구에서의 분류는 결함 유형 속성은 간단하나, 결함 데이터에 결함 트리거 속성(결함을 드러나게 하는 조건)도 포함된다는 특징이 있다.

• **Eisenstadt의 버그 스토리 분류 (Eisenstadt, 1997)**

- Eisenstadt은 프로그래밍 오류의 다양한 측면을 자세히 설명하기 위해 프로그래머들에게 그 오류가 왜 어려운가, 어떻게 발견했는가, 근본 원인은 무엇인가와 같은 세 가지 질문을 가지고 버그 보고서를 요청했고, 프로그래머들이 제공한 59개의 버그 리포트를 분석하였다.
- 결함의 근본 원인에 대한 질문에 대한 답은 SW 결함에 대한 또 다른 분류로 이어질 수 있다.
  - ✓ mem - 의도치 않게 덮어 써지거나 소모된 메모리 (13)
  - ✓ 공급업체 - 공급업체의 문제(하드웨어 또는 SW) (9)
  - ✓ des.logic - 예상치 못한 경우(잘못된 설계 논리) (7)
  - ✓ init - 잘못된 초기화; 잘못된 타입; 정의 충돌 (6)
  - ✓ lex - 어휘 문제, 잘못된 구문 분석 또는 모호한 구문(4)
  - ✓ var - 잘못된 변수 또는 연산자 (3)
  - ✓ 미해결 - 현재까지 알려지지 않은 미해결 (3)
  - ✓ lang - 의미가 모호하거나 잘못 이해됨(2)
  - ✓ behav - 최종 사용자(또는 프로그래머)의 파악하기 어려운 행위(2)
- 이 분류에서는 프로그래밍 오류 뿐 아니라 최종 사용자의 잘못된 행동과 타사 하드웨어 또는 SW 문제도 고려되었다.
- 이 연구에서 분류의 목적은 다른 사람들이 발견하고 보고한 오류들을 그룹 짓는 것이었다. 결함 스토리로부터 배울 점들에 초점을 두었다. 이 연구는 향후 디버깅 도구들이 어떻게 설계되어야 할지에 대한 조언을 주려고 노력하였다.

• **DeMillo 및 Mathur의 문법 기반 결함 분류 스키마 (DeMillo and Mathur, 1995)**

- 디버깅을 지원하기 위한 목적으로 결함을 분류하였다.
- 결함을 “프로그래밍 실수가 드러난 것“으로 정의하였고, 결함이 소스 코드에서

삽입되거나 제거되는 중요한 문자열로 식별할 수 있다고 여겼다.

- 순수하게 문법적 접근 방식을 취하여 정형적으로 개발된 문법 기반 분류 스키마를 제시하였다. 분류 스키마는 네 가지 주요 범주가 있는 계층 구조로 구성된다.
  - ✓ 가짜 엔터티 – 특정 부분 문자열 제거 필요 (2/0.5%)
  - ✓ 누락된 엔터티 – 문법적 엔터티의 삽입이 필요 (155/53%).
  - ✓ 잘못 배치된 엔터티 – 위치 변경이 필요함 (16/5.5%).
  - ✓ 잘못된 엔터티 – 기타 다른 결함들 (118/41%)
- 저자들은 문법(syntax)을 의미(semantics)를 전달하는 매개체로 간주하였다.
- 분류의 목적은 수동 분류에 오래 걸리는 시간도 줄이고, 모호성도 줄이려고 SW 결함 분류를 자동화하려는 것이었다.
- 상당 부분 오류가 초기 4개월간 테스팅에서 발견된 반면, 나머지는 이후 5년간의 테스팅에서도 발견되지 않았던 사례를 바탕으로, 결함의 빈도 이외에도 결합 타입의 지속성(persistence)이 중요한 결함 특성이라고 결론지었다.

- **Du와 Mathur의 보안 오류 분석 기반 분류 스키마 (Du and Mathur, 1998)**
  - Du와 Mathur는 보안 침해로 이어진 일련의 오류를 분석하고 이러한 오류를 분류하는 새로운 스키마를 개발하였다. 전반적인 목표는 중요한 SW 문제를 드러낼 때 다양한 코드 커버리지 측정의 효율성을 평가하는 것이다.
  - 분류는 세 가지 관점으로 구성된다. 첫 번째 관점은 보안 오류의 원인(예, 유효성 검사 오류, 인증 오류)을 지정하고, 두 번째 관점은 오류의 영향(예, 코드의 허가되지 않은 실행 또는 서비스 거부)을 나타낸다. 세 번째 관점은 오류 수정을 지정하고 DeMillo 및 Mathur(DeMillo and Mathur, 1998)의 스키마와 동일한 4가지 속성으로 구성된다.
  - 분류가 여러 관점으로 이루어진다는 점을 참고할 수 있다.

Research Issues in Software Fault Categorization 연구에서는 여러 기존 연구들에서 SW 결함을 어떻게 분류하였는지 분석하였다. 각각 다른 저자들이 목적에 따라 다른 결함 카테고리를 제안한 것을 확인 하였으며, 분류 기준이 적합한지에 대한 충분한 설명이 제시되지 않은 사례도 많았다. SW 결함 관련한 정량적 관찰 결과를 일반화하거나 다른 프로젝트로 전이하는 것은 여전히 열린 이슈임을 시사하였다.

2. SW 실패를 고려한 연구

기존의 SW 결함 분류 연구들은 SW 결함 유형 분류가 목적에 따라 어떻게 달라질 수 있는지 볼 수 있으나, 이 연구들은 위험 분석을 목적으로 분류한 것이 아니어서 본 연구의 목적인 사고 위험원 분석을 위한 모델 개발에 참조하기에 한계가 있다. 이에 위험 분석에서 SW 실패를 고려한 연구들을 조사 분석하였다.

Thieme 외 연구팀(Thieme et al., 2020)은 “Incorporating software failure in risk analysis-Part 1: Software functional failure mode classification” 연구에서 SW 기능(function)과 관련된 기능적 실패 모드(functional failure mode)에 대한 개념을 정의하고, SW ‘기능’에 대한 기능적-실패 모드 분류법(taxonomy)을 제안하였다. 위험 분석에서 하드웨어, SW, 인간, 조직의 위험 요소들이 충분히 다루어져야 하나, 기존 위험 분석에서 SW 실패에 대한 부분이 종종 간과되었다. 이 연구에서는 일반적인 위험 분석 프레임워크에 SW 실패를 포함할 수 있는 기반을 제공하고자 하였다.

기존 SW 실패 모드 분류 연구 조사 분석을 위해 IEEE(Institute of Electrical and Electronics Engineers) 데이터베이스 Xplore 및 Scopus에서 아래의 키워드들을 사용하여 2000년도 이후 출판물 대상 검색을 수행하였다[8].

- software failure mode identification(SW 실패 모드 식별)
- software FMEA(SW FMEA)
- software FMECA(SW FMECA)
- software failure mode effect analysis(SW 실패 모드 효과 분석)
- software failure mode effect criticality analysis(SW 실패 모드 영향 심각성 분석)
- 관련 분류를 포함하는 출판물을 면밀히 조사하고 선별된 건에 대해서는 추가 분석을 하였다. 관련 분류 체계가 2002년부터 2014년 사이 출판물에서 발견되었다.

Li 외 연구팀(Li, 2004; Li et al., 2003)은 SW 기능에 대한 실패 모드 분류 체계를 개발하였다. 1) 기능적-실패 모드(속성 및 기능 또는 기능-집합 실패 모드) 및 2) 외부적 실패 모드(타이밍, 입력/출력, 다중 상호작용, 지원(support) 실패 모드)의 두 가지 범주에서 여러 유형의 실패 모드를 정의하였다.

8) 최근 연구 출판물에 이전 분류 체계가 반영되었다고 가정했다.

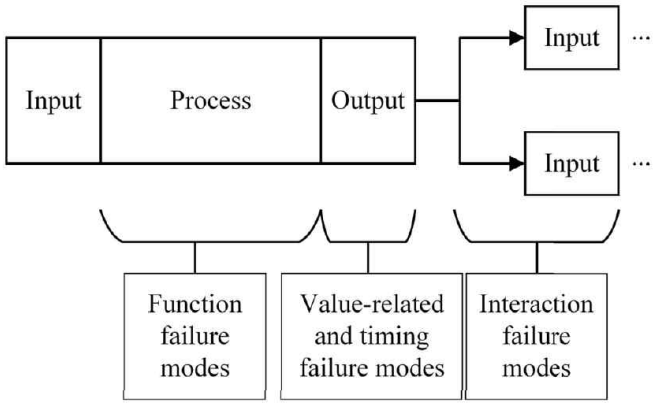
- 입력/출력 실패 모드는 SW 기능 자체에서 기인하지 않는 모드이다(Li et al., 2003). 입력 실패는 출력 실패로 이어진다. 입/출력 실패 모드는 값 관련 값-관련 실패 모드와 타이밍 관련 실패 모드로 더 나눌 수 있다(Li et al., 2003).
- 다중 상호작용 실패 모드는 정보를 교환하기 위해 공통 언어를 통해 통신하는 것과 관련이 있다(Li et al., 2003). 지원 실패 모드는 하드웨어 리소스 및 물리적 운영 환경과 관련된 모드로 구성되므로 SW 기능에 직접 적용되지 않는다. SW 실패 원인 범주에 속하지만 이 연구에서는 자세히 다루지 않았다.

경제협력개발기구(OECD) 원자력에너지부(OECD, 2014)는 Li 외 연구팀(Li et al., 2003; OECD, 2014; Li et al., 2005)과 Authen 및 Holmberg 연구팀(Authen et al., 2010; Authen and Holmberg, 2012; Authen and Holmberg, 2013; Holmberg et al., 2012) 연구에 기반하여 하드웨어 및 SW 실패 모드 분류 체계를 제안하였다. 전체 시스템, 부서, 계측 및 제어 수준과 같이 시스템의 다양한 수준을 다룬다.

Ristord와 Esmenjaud(Ristord and Esmenjaud, 2002), Huang 외 연구팀(Huang et al., 2009), Stadler와 Seidl(Stadler and Seidl, 2013), Park 외 연구팀(Park et al., 2014), Prasanna 외 연구팀은(Prasanna et al., 2014) SW FMEA를 자체적으로 개작한 것을 제안하고 논의하였다.

Thieme 외 연구팀은 이전 연구들이 가능한 실패 모드 식별을 위한 기반을 제공하지만 분류체계에 대한 명확한 설명이나 대상으로 하는 SW 추상화 수준(예, SW 시스템, SW 기능, 코드 분석) 간의 명확한 구분이 없었던 점을 보완하고자, 위험 분석 목적으로 SW 기능과 그와 관련된 SW 시스템 실패 모드에 대한 개념을 정의하고 명확히 하였다. SW 실패로 인한 위험에 초점을 두었으며, SW에서의 기능적 관점을 이용하여, 실패 모드(failure mode), 실패 원인(failure cause), 실패 영향(failure effects)을 구분하여 다루었다.

Thieme 외 연구팀은 기능(function), 상호작용(interaction), 타이밍 관련(timing-related), 값 관련(value-related)와 같은 네 가지 카테고리에 대한 기능적 실패 모드 분류를 제시하였다. [그림 3-2]는 SW 기능과 모든 수준의 기반이 되는 컴포넌트에 대한 단순화된 뷰로 Huang 외 연구(Huang et al., 2009)에서 제안된 것을 확장한 것이다. 대상 시스템을 기능 분해(functional decomposition)를 거쳐 적절한 수준으로 분해하였을 때, 기능(function) 및 주요 요소를 나타낸 것이다. 다양한 범주의 실패 모드를 적용할 수 있는 위치가 표시되어 있다.



[그림 3-2] SW 기능과 모든 수준의 기반 컴포넌트

- **기능 실패 모드** : 프로세스(Process) 섹션은 기능적 동작과 계산이 실행되어 입력을 출력으로 바꾸는 곳이다. 기능 실패 모드는 이 입력-프로세스-출력 섹션과 연결된다. <표 3-2>는 SW-기능 실패 모드 분류 체계이다.

<표 3-2> SW-기능 실패 모드 분류 체계

실패 모드	부가 설명
기능 누락 또는 누락된 작업	기능(또는 그 일부분)이 실행되지 않음
부정확한 기능	기능이 의도한 작업을 실행하지 않음
부가적인 기능	기능에서 부가적이고 명세 되지 않은 작업이 실행됨
투표 없음	기능 안에서 입력에 대한 투표가 수행되지 않음
부정확한 투표	입력에 대한 투표가 기능 내 명세에 따라 수행되지 않음. 따라서, 투표 결과가 부정확함
실패 처리에 실패	탐지된 실패가 적절하게 처리되지 않음

- **값 관련 / 타이밍 관련 실패 모드** : 기능은 항상 하나 이상의 출력을 가진다. 출력은 수치적인 값, 이진 값, 또는 특정 기능 호출일 수도 있다. 입력은 또 다른 기능의 출력이며 외부 인터페이스를 통해 제공된다. 한 기능의 출력은 여러 다른 기능의 입력이 될 수 있다. 하나의 기능이 여러 개의 입력 및 여러 개의 출력을 가질 수 있다. 입출력은 관련된 데이터 타입과 허용범위를 가진다. 출력이 데이터 배열 또는 구조의 일부인 경우 데이터 형식은 요소의 순서를 나타낸다. 데이터 속도 및 버퍼(데이터 또는 이벤트를 수집하기 위해 입력 또는 출력에 사용되는 타입)는 해당되는 경우에만 기술하면 된다. 값-관련 및 타이밍-관련 실패 모드가 SW 기능의 이 부분과 연계된다. <표 3-3>은 SW 기능의

타이밍 관련 실패 모드에 대한 분류이며, <표 3-4>는 SW 기능의 값 관련 실패 모드에 대한 분류이다.

<표 3-3> SW 기능의 타이밍 관련 실패 모드에 대한 분류

실패 모드	상세 실패 모드	추가 설명
출력 제공	너무 이른	출력이 요구되지 않거나 필요하지 않은 때 제공됨 기능으로부터 출력이 제공되지 않음
	너무 늦은	
	가짜의	
	순서가 맞지 않는	
	제시간이 아닌	
출력률 실패	너무 빠른	시간 출력의 길이가 가용함
	너무 늦은	
	일관되지 않은	
	동기화되지 않은	
지속	너무 긴	
	너무 짧은	
반복 기능이 부정확하게 스케줄됨		주기적으로 요구되는 출력이 기대한 시간에 도달하지 않음

<표 3-4> SW 기능의 값 관련 실패 모드에 대한 분류

실패 모드	상세 실패 모드	추가 설명
값이 없음		값이 제공되지 않음
잘못된 값	너무 높음	값이 예상 또는 필수 값보다 높음. 1이거나 최대 허용되는 값 또는 더 높은 값 증가일 수 있음
	너무 낮음	값이 예상 또는 필수 값보다 낮음
	반대값	값이 예상 값의 반대값임
	값이 0임	값이 예상 값 대신 0임
범위를 벗어난 값	데이터 타입 허용 범위	
중복/동결된 값	응용프로그램이 허용한 범위	지속적으로 동일한 값이 생성됨
노이즈 값/정확도 오류		전송되는 값이 충분히 정확하지 않음
잘못된 데이터 타입을 가진 값		

숫자가 아닌 값	숫자가 아님(NaN)	SW에서 해석할 수 없는 값이 전송됨
	무한(infinite)	
	음의 무한(negative infinite)	
데이터 배열이나 구조에 있는 원소	너무 많음	
	너무 적음	
	잘못된 순서	
	거꾸로 된 순서	
올바른 값이 잘못된 것으로 확인됨	열거된 값이 잘못됨	데이터 배열 또는 구조에서 잘못된 요소가 참조됨
잘못된 값이 올바른 값으로 확인됨		
데이터가 검증되지 않음		유효성 검사가 실행되지 않음

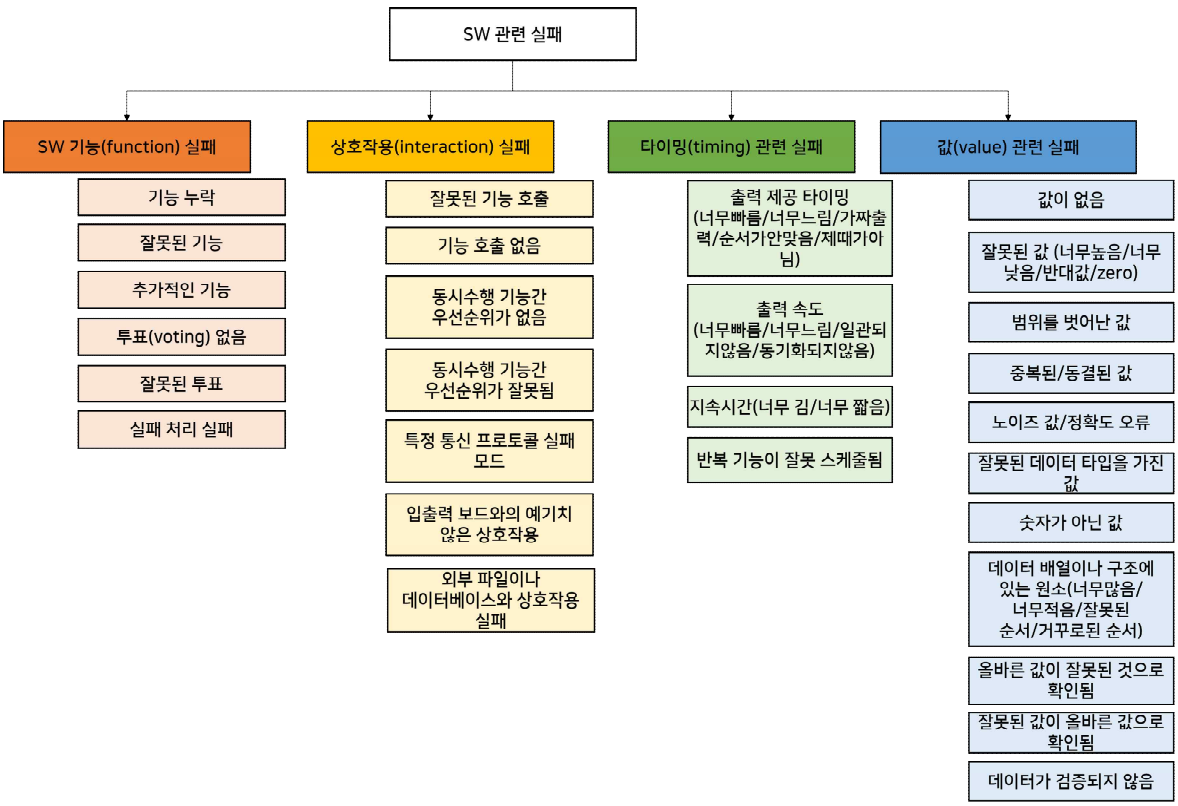
- **상호작용 실패 모드**: SW 시스템에서 기능들은 특정 순서대로 실행된다. 주기적으로 실행되기도 하고 요청 시 실행되기도 한다. 각 기능은 정보를 전달하거나 다른 기능을 호출한다. 이러한 상호작용은 상호작용 실패 모드와 연계된다. (SW와 상호작용하는) 외부 인터페이스는 다른 SW 시스템, 센서, 데이터베이스, 운용하는 사람 등을 포함한다.

<표 3-5> SW 기능의 상호작용 실패 모드에 대한 분류

실패 모드	상세 실패 모드	추가 설명
우회되거나 부정확한 기능 호출		현재 기능이 종료된 후 잘못된 기능이 호출됨
다음 기능에 대한 호출이 없음		현재 기능이 완료된 후 더이상 기능이 호출되지 않음
동시 수행 기능에 대한 우선순위 없음		동시에 실행해야 하는 기능들에 대한 호출에 우선순위가 부여되지 않음
동시 수행 기능에 대한 잘못된 우선순위		동시에 실행해야 하는 기능들에 잘못된 우선순위가 부여됨
통신 프로토콜-종속적인 실패 모드		SW 시스템의 일부 간에 정보를 교환하는데 사용되는 특정 통신 프로토콜에 고유한 실패 모드

입출력(IO) 보드와의 예기치 않은 상호작용		입출력 보드 또는 인터페이스와의 상호작용(가짜일 수 있음)과 관련된 실패 모드
외부 파일 또는 데이터베이스와의 상호작용 실패	잘못된 이름	파일 또는 데이터베이스 이름이 올바르지 않음
	유효하지 않은 이름/확장자	파일 또는 데이터베이스에 대해 입력한 이름에 유효하지 않은 기호가 있음
	파일/데이터베이스가 존재하지 않음	파일 또는 데이터베이스가 정확하게 명세되어 있는 것처럼 보이지만 존재하지 않음
	파일/데이터베이스가 열려 있음	파일 또는 데이터베이스가 다른 프로그램에서 열려 있고 다시 열 수 없음
	잘못된/유효하지 않은 파일 형식	파일 형식이 예상 파일 형식과 다름
	파일 헤더에 오류가 있음	파일 헤더 정보에 필요한 정보와 다른 정보가 있음
	파일 끝에 오류가 있음	파일 끝 정보에 필요한 정보와 다른 정보가 있음
	잘못된 파일 길이	파일 길이가 요구되는 또는 예상된 길이와 다름
	파일/데이터베이스가 비어 있음	
	잘못된 파일/데이터베이스 내용물	파일 또는 데이터베이스의 내용이 예상된 또는 요구되는 정보와 다름

Thieme 외 연구에서는 29개의 실패 모드를 식별하여 분류 체계를 구성하고, 사례연구를 통해 실증하였다. [그림 3-3]은 식별된 실패 모드를 시각화 한 것이다. 사례 연구에서 대상 SW 프로그램에 제안한 실패 모드를 적용하였고, 분류법을 적용함으로 명확한 실패 모드 파악으로 이어질 수 있음을 보였으며, SW 집약적 시스템의 위험 분석에 입력이 되는, SW 오류 모드를 식별하는 데 제안하는 분류 체계가 도움이 됨을 보였다.



[그림 3-3] Thieme 외 연구에서 식별한 SW 실패 모드 도식화

<표 3-6> SW 기능 수준에 관련된 실패 모드 유형 연구들

주저자		Ristord and Esmenjaud (Ristord and Esmenjaud, 2002)	Li (Li, 2004)	Wei	Huang et al. (Huang et al., 2009)	Stadler and Seidl (Stadler and Seidl, 2013)	OECD (OECD, 2014)	Park et al. (Park et al., 2014)	Prasanna et al. (Prasanna et al., 2014)	Thieme (Thieme et al., 2020)
분류 목적		위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	**위험 분석**
제시된 실패 모드 개수		12	31	12	25	21	37	21	11	**29**
실패 모드 커버리지	기능	Yes	Yes	No	No	Yes	Yes	Yes	No	**Yes**
	상호작용	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	**Yes**
	타이밍 관련	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	**Yes**
	값 관련	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	**Yes**

<표 3-6>은 SW 기능 수준에 관련된 실패 모드 유형을 포함한 연구들의 요약이다. Thieme 외(Thieme et al., 2020) 연구가 가장 최근 연구이면서, 제안한 네 가지 카테고리(기능, 상호작용, 타이밍 관련, 값 관련) 및 세부 29가지 실패모드가 기존 연구들에서 제안한 실패 모드들을 포괄하기에 Thieme 외 연구를 SW 시스템 위험원 분석 모델의 기반으로 선정하였다.

3. ICAM 사고 원인분석 방법론

Thieme 외 연구에서 제안된 SW 실패 모드 분류 모델은 'SW 기능 수준'에 관련된 위험원 분류 모델로는 가장 적합하나, SW 시스템 경계를 기준으로 내부만을 대상으로 하기 때문에, SW 시스템과 관련이 있으나 위험원이 SW 외부에 있는 경우를 포괄하지 못하는 한계가 있다. 이에 SW 시스템 경계 외부의 액터, 조직, 환경, 하드웨어 등까지도 포괄할 수 있는 위험원 모델을 고려하였다.

Incident Report의 전문가 의견 작성은 전문가 다수를 통한 In-depth 인터뷰 접근을 통한 의견 조율을 진행하였다. 사고 원인에 대한 전문가 의견을 체계적으로 작성하여 논의에 들어가기 위한 방법으로 ICAM(Incident Cause Analysis Method)(Landre et al., 2006)를 활용하였다.

Incident Cause Analysis Method는 사고분석 전문가인 영국 맨체스터 대학교의 James Reason 교수에 의해 제안된 방법론으로 인위적으로 예방 가능한 사고 원인들을 분석하는 데 사용되는 방법론이다. 이 방법론은 간단하나 특히 크고 복잡한 사고, 사회에서 사용되는 기술 사고(예, 항공, 광산 등)에 효과적으로 접근하는 방법론이다.

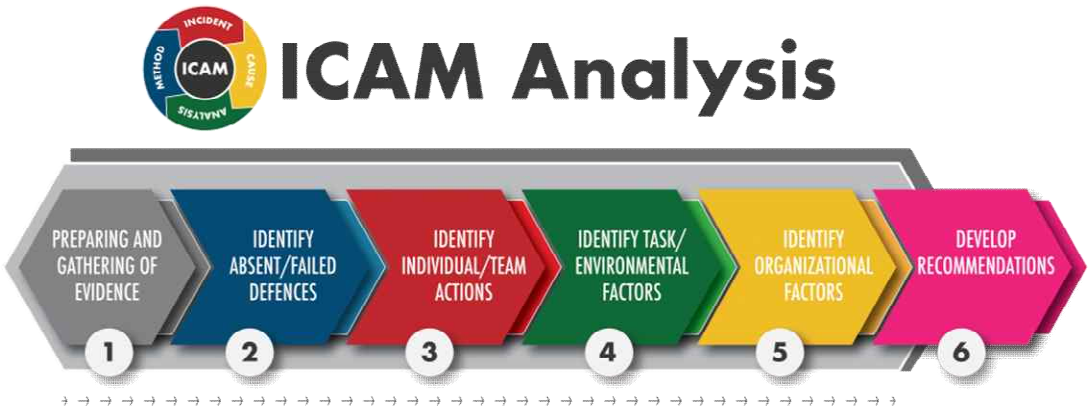
Reason 교수는 사고의 발생 원인을 크게 다음과 같은 복합적인 관점에서 바라본다. Reason 교수는 사고는 latent condition(잠재되어 있는 사고 유발 조건으로서, 경영/관리적 의사결정, 통상적 실무, 문화적 영향을 포함), trigger condition, 그리고 active failure의 복합체라고 설명한다. 이 latent condition에 잠재된 부정적인 요인들이 실제 사고 당시의 triggering condition(기후, 지역 등)과 결합 되고 또한 active failures (개인이나 팀에 의해 유발된 오류 혹은 절차적 결함)와 결부되어 사고를 발생한다고 보고 있다.

ICAM의 기본적인 개념은 또한 인간의 오류를 필연적으로 받아들이고 있다는 것이다. 인간의 오류(error)는 인간 행동학 측면에서 일반적으로 일어날 수 밖에 없는 것이기 때문이다. 물론 인간의 오류를 줄일 수는 있으나 완전히 제거할 수는 없다. ICAM은 이러한 인간의 오류를 감안한 조사 방법론으로서 사용자와 SW로 구성된 SW 시스템에 대한 조사 관점에서 매우 유용하다고 판단된다.

ICAM을 사용하는 용도들은 주로 다음과 같다.

1) 어떤 사건에 대하여 포괄적인 자료 및 사실 확립을 위해서
2) 조사 자체가 사용자 혹은 운영자의 실수에만 한계지어 지는 것을 방지하기 위해서
3) 사고의 잠재 원인들을 파악하기 위해서
4) 현재 존재하는 혹은 실행 중인 제어나 절차의 적절성 판별을 위해
5) 차후 사고 재발 방지를 위한 corrective action 제안을 위해(위험 감소, 재발 방지)
6) 발생/재발하고 있는 사고의 동향을 파악하기 위해
7) 사고조사의 목적이 특정 사람이나 조직을 비판하는 것이 아니라 객관적 요인을 찾고 방지책 개발을 위해
8) 사고조사에 필요한 법적 근거들을 마련하기 위해

[그림 3-4]는 ICAM의 절차이다.



[그림 3-4] ICAM 절차

ICAM에서는 사고 요인들을 다음과 같이 4가지의 요소로 정리한다.

- Absent / Failed Defense
  - 이 요인은 시스템의 기술적 failure 혹은 사용자의 failure에 대한 부적절한 대응 방어책이 부재들을 의미한다. SW 공학 관점에서 보았을 때 통상적으로 Safety Mechanism의 부재라고 볼 수 있다.

- Individual / Team Actions
  - 이 요인은 사고를 유발한 사용자, 운영자, 타 시스템들과 관련된 에러, 위법적 행위들을 의미한다. SW 공학적 관점에서 보았을 때는 system context diagram 에서 actor에 해당하는 entity의 오류들을 의미한다고 볼 수 있다.

- Task / Environmental Condition
  - 이 요인은 사고 발생 직전 혹은 사고 발생 당시의 환경적 요인들을 의미한다. 통상적으로 시스템의 운영 환경, 또한 사용자/운영자의 작업환경의 사고 당시 요인들을 말한다. SW 공학적 관점에서 볼 때 사고 당시의 시스템 혹은 액터가 처한 환경으로 볼 수 있다.

- Organizational Factors
  - 이 요인은 실제 운영 혹은 작업환경에 영향을 미치는 조직적 요인들을 의미한다. 조직적인 요인들은 매우 광범위할 수 있는데, ICAM에서는 다음과 같이 14개의 organizational factor type을 제시하고 있다.

HW	Hardware
TR	Training
OR	Organisation
CO	Communication
IG	Incompatible Goals
PR	Procedures
MM	Maintenance Management
DE	Design
RM	Risk Management
MC	Management of Change
CM	Contractor Management
OC	Organisational Culture
RI	Regulatory Influence
OL	Organisational Learning

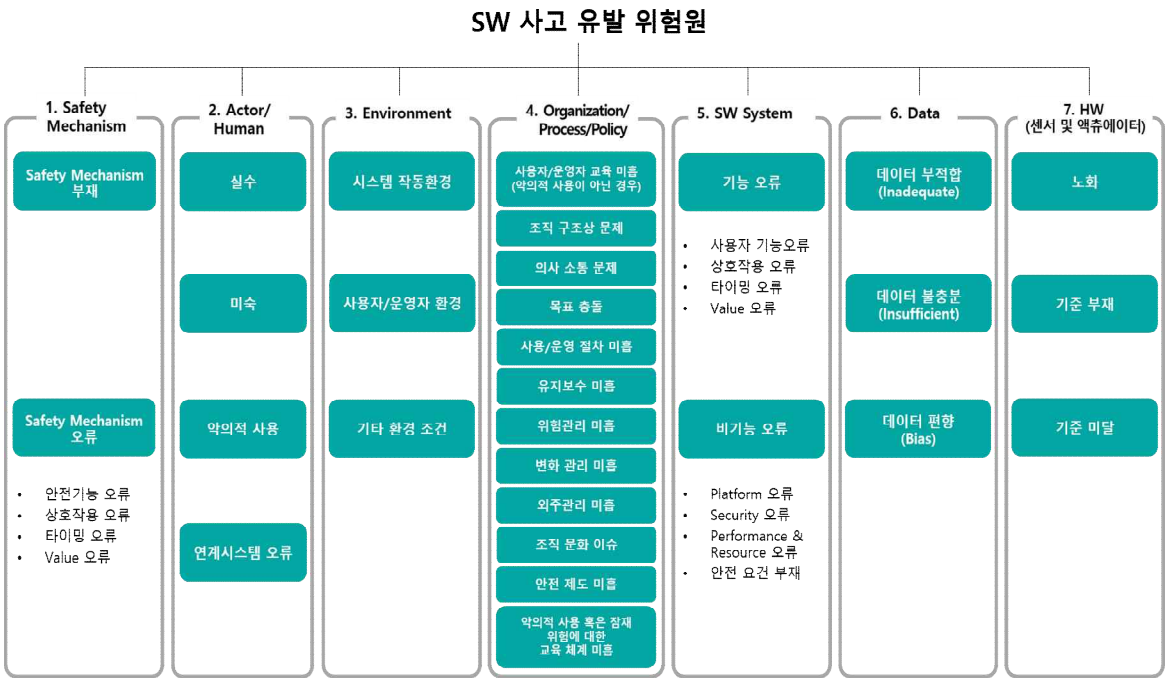
[그림 3-5] ICAM에서 제시한 14가지 organizational factor type

제3절 SW 신기술 사고 유발 위험원 유형화

본 연구에서는 ICAM 방법론을 SW시스템의 특성에 맞게 개선하여 사용하고자 한다. 개선된 방법론을 편의상 'SW-ICAM' 로 명명하도록 한다. SW-ICAM은 ICAM에서 제안하는 4가지 관점을 소프트웨어 공학에서 다루는 SW오류 분류 체계들에 근거하여 7가지로 세분화하여 개발하였다.

[그림 3-6]은 SW-ICAM에서 제안하는 7가지 관점의 SW사고 유발 위험원 프레임워크이다. 이 7가지 분류는 ICAM에서 제안하는 4가지 관점을 모두 포함하고 있으며, 그 외 SW 특면에서 보강 및 세분화 되어 있다.

ICAM에서는 조직 분야에만 세부 요인 유형들을 제시하고 있는데 SW-ICAM 에서는 7개의 모든 분야에 대하여 세부적인 요인 유형들을 제공하였다. 다음은 각 분야에 대한 설명이다.



[그림 3-6] SW 신기술 사고 유발 위험원 유형화

(1) **Safety Mechanism 부재 혹은 오작동** : SW 시스템에서의 Safety Mechanism에 해당하는 것으로 통상적으로 예측 가능한 잠재 위험이 실제로 발생하였을 때 SW 안전 기능을 통하여 사고를 방지 혹은 최소화하는 기능을 의미한다. 이는 기능 안전(Functional Safety)분야에서 국제 표준으로 요구되는 기능인데, 기능 안전이

자동차 전장 부품과 같은 분야에는 존재하나 아직 SW신기술 분야에서는 표준화가 되어 있지 않다. 표준화가 되어있지 않다고 불필요한 것으로 여기는 것은 아니며, 다만 SW신기술 분야이다 보니 아직 안전 표준이 존재하지 않는 것으로 판단하면 된다. Safety Mechanism의 세부 요인은 다음과 같다.

Safety Mechanism 부재: SW 안전 기능 부재	SW Safety 메커니즘 자체가 없는 경우
Safety Mechanism 오류: 안전 기능 오류	SW Safety 메커니즘은 있으나 기능이 오작동한 경우
Safety Mechanism 오류: 상호 작용 오류	오류의 원인이 상호작용에 있는 경우
Safety Mechanism 오류: 타이밍 오류	오류의 원인이 실시간 타이밍 오류에 있는 경우
Safety Mechanism 오류: Value 오류	오류의 원인이 상수 혹은 변수값 오류에 기인한 경우

(2) Actors/Human 오동작 : SW 시스템에서 사용자 혹은 운영자의 잘못된 조작을 의미한다. 참고로, SW 안전 분야와 제조물책임법에서는 'Reasonable Foreseeable Misuse' 라는 표현을 사용하는데, 이는 "use of a product or system in a way not intended by the manufacturer, but which can result from readily predictable human behavior" 로 해석된다. 안전성을 요하는 시스템에서는 이러한 Reasonable Foreseeable Misuse에 대하여 사전에 대비하여 안전 메커니즘을 구현해야 한다. 이번 연구 분석에서의 Actor 오동작은 안전 메커니즘의 유무를 떠난 Actor의 오동작을 서술한다. 다만 해당 오동작이 Reasonable Foreseeable Misuse 인지 혹은 전혀 예측할 수 없었던 처음 일어난 오동작인지에 대한 전문가 의견을 다룰 수 있다. 추가적으로 여기서 Actor는 사람일 수도 있고 혹은 해당 시스템에 연결된 타 시스템일 수도 있다. Actors/Human의 세부 요인는 다음과 같다.

Human 실수	사용자 혹은 운영자의 조작 실수
Human 미숙	사용자 혹은 운영자의 조작 미숙
Human 악의적 사용	사용자 혹은 운영자의 악의적 사용
연계(Peer)시스템 오류	연계시스템의 오류로 인한 문제 파급

(3) Environmental 요인 : 시스템이 작동하는 환경적 요인, 혹은 사용자/운영자가 조작하는 환경적 요인을 의미한다. 통상적으로 자율주행자동차 SW와 같은 경우 주행 당시의 기후에 민감하다. SW신기술의 환경에 대하여서는 본 연구에서 실험적으로 확인이 필요한 분야이다. Environment 분야의 세부 요인는 다음과 같다.

시스템 작동 환경	SW 시스템의 작동 환경을 의미한다. 여기서는 SW가 내장된 시스템의 작동환경을 포함한다. 또한 네트워크 상의 환경도 포함된다.
사용자/운영자 작업 환경	사용자 및 운영자의 작업 환경으로서 SW의 경우 작업 장소에 관련된 부분이다.
기타 환경 조건	기타 환경적 조건을 명시하여야 하는 부분

(4) Organization/Process/Policy 요인 : ICAM에서의 Organizational 요인에 해당하는 분류는 SW 분야의 특성에 맞게 재해석할 필요가 있다. SW-ICAM에서는 ICAM에서의 organizational 요인을 크게 4가지로 재분류 및 보강, 변경하였다. 첫째로 Organization/Process/Policy이고 그리고 SW System, Data, HW 순으로 재분류하였다. 먼저 Organization/Process/Policy는 SW 개발/운영 조직상의 문제, SW 개발 및 운영 절차, 교육, 개발 과정 등을 다루고 있다. 다음은 Organization/Process/Policy의 세부 요인들이다.

사용자/운영자 교육 미흡 (악의적 사용이 아닌 경우)	특정 시스템에 대한 사용자 혹은 운영자 교육이 미흡한 경우이다.
조직 구조상 문제 혹은 이슈	조직 구조상의 문제로서 예를 들면 역할 및 책임이 불분명하여 생긴 문제들이다.
의사소통 문제	시스템과 관련된 관계자(Stakeholder) 간의 의사소통 문제이다.
목표 충돌 문제	시스템을 운영 혹은 사용하는데 있어서 두 가지 이상의 목표가 충돌한 경우이다.
사용/운영 절차 미흡	사용자 매뉴얼, 혹은 절차서 등 업무 지침이 미흡한 경우이다.
유지 보수 미흡	SW의 필드 결함 관리, 업데이트 등 유지보수와 관련된 내용이다.
위험 관리 미흡	SW 개발 시 잠재 위험에 대한 분석이 미흡하여 발생한 경우이다.
변화 관리 미흡	사용자, 운영자 조직의 혹은 사회의 변화관리가 미흡한 경우를 의미한다. 예를 들자면 메타버스 같은 신기술의 경우 사회 차원에서 적절한 변화 관리가 필요하다.
외주 관리 미흡	SW 개발 시 활용한 외주업체에 대한 관리(스펙, 계약 등)상의 문제를 의미한다.
조직 문화 이슈	안전 불감증과 같은 조직 차원의 문화적 이슈들을 의미한다.
안전 제도 미흡	해당 SW의 안전한 사용을 위한 필요한 규정 혹은 제도가 미흡한 경우를 의미한다.
악의적 사용 혹은 잠재 위험에 대한 교육체계 미흡	악의적 사고 혹은 기존에 없었던 잠재 위험에 대한 교육체계가 미흡한 경우이다.

(5) SW System 요인 : SW System의 요인은 SW 공학적인 관점에서 해당 SW에서 일어나는 오류들을 의미한다. 이러한 오류들은 크게 사용자 기능(Functional)에 대한 오류와 비기능적(Non-Functional) 오류로 분류할 수 있다. 기능 오류와 비기능 오류는 특성에 따라 다음과 같이 세분화 된다. 다음은 SW 시스템의 세부 요인들이다.

기능 오류: 사용자 기능 오류	통상적인 기능 구현상 오류이다. 로직 오류, 알고리즘, 구조 등이다.
기능 오류: 상호작용 오류	타 시스템과의 상호작용에 오류가 생겨 기능 오류가 발생한 경우이다.
기능 오류: 타이밍 오류	실시간 시스템에서 타이밍상의 오류가 기능 오류를 야기시키는 경우를 말한다.
기능 오류: Value 오류	기능에서 사용하는 값에 오류가 있어 오작동한 경우를 의미한다.
비기능 오류: Platform 오류	플랫폼과의 호환 상의 오류를 의미한다. 운영체계 호환 오류 등이 여기 속한다.
비기능 오류: Security 오류	SW 보안상의 오류가 사고를 유발한 경우를 의미한다.
비기능 오류: Performance & Resource오류	SW 성능과 자원(메모리, CPU 등)상의 이슈가 오류로 발전한 경우이다.
비기능 오류: 안전 요건 부재	해당 SW가 필요로 하는 안전 요건에 대한 파악이 미흡하거나 요건 자체가 부재한 경우이다.

(6) Data 요인 : 머신러닝 등 대규모 데이터를 활용하는 SW가 급증함에 따라 데이터와 관련된 오류들 별도로 정리하였다. 다음은 데이터 분야의 세부 요인들이다.

데이터 부적합 (Inadequate) (부적절 획득 포함)	잘못된 데이터셋을 사용하는 경우를 의미한다. AI, 빅데이터 분석 등에 부적합한 데이터를 학습 혹은 분석하는 경우를 의미한다. 또한 데이터를 부적절하게 획득한 경우도 포함한다.
데이터 불충분 (Insufficient)	불충분한 데이터셋을 사용하는 경우를 의미한다. AI 등에서 불충분한 데이터를 기반으로 학습한 경우가 이에 속한다.
데이터 편향 (Bias)	AI에서 편향된 데이터를 학습하여 생기는 문제들을 의미한다.

(7) HW 요인(센서 혹은 액츄에이터) : 마지막으로 SW가 연결된 해당 시스템의 하드웨어 상 문제들을 다룬다. 기본적으로 HW가 오동작하여 생기는 문제들이 이에 해당한다. HW 분야의 세부 요인은 다음과 같다.

노화	센서 등 하드웨어가 시간에 지남에 따라 노화되어 오동작하는 경우를 의미한다.
기준 부재	하드웨어의 필요 성능, 품질에 대한 기준이 부재한 경우를 의미한다.
기준 미달	하드웨어의 필요 성능, 품질에 대한 기준은 존재하나 이에 미달하는 부적합한 하드웨어를 의미한다.

<표 3-7>은 SW 시스템 사고 위험원 분석과 관련된 연구들의 위험원 커버리지를 비교 분석한 결과이다. 분류 목적에 따라 다양한 종류의 분류 체계/모델이 제안되었고, 위험원 커버지리를 보았을 때, SW-ICAM 모델이 기존 분류 체계에 비해 SW 시스템 내부 위험원 뿐만 아니라, Safety Mechanism, Actor/Human, Environment, Organization/Process/Policy, Data, HW를 포함하는 포괄적인 분류 체계를 제시하고 있다.

<표 3-7> 관련 연구들의 위험원 커버리지 비교 분석

			SW 오류 분류 연구들						SW 기능(function) 수준에 관련된 실패 모드 유형을 포함한 연구들									사고 원인 분석
연구진			Knuth	Beizer	Gray	Chillarege	Eisenstadt	DeMillo	Ristord and Esmenjaud	Li	Wei	Huang	Stadler and Seidl	OECD	Park	Prasanna	Thieme	SW-ICAM
분류 목적			교육	테스팅	결함 감내	프로세스 개선	디버깅	테스팅	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	위험 분석	**위험 분석**
카테고리 (제시된 failure mode) 개수			9	>100	2	8+18	10	>20	12	31	12	25	21	37	21	11	29	**36**
위험원 커버리지	SW 시스템	기능	NA	NA	NA	NA	NA	NA	O	O	X	X	O	O	O	X	O	**O**
		상호 작용	NA	NA	NA	NA	NA	NA	O	O	X	O	O	O	O	O	O	**O**
		타이밍 관련	NA	NA	NA	NA	NA	NA	O	O	O	O	O	O	O	X	O	**O**
		값 관련	NA	NA	NA	NA	NA	NA	O	O	O	O	O	O	O	O	O	**O**
	Safety Mechanism		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	**O**
	Actor/ Human		△	△	X	△	X	X	X	X	X	X	X	X	X	X	X	**O**
	Environment		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	**O**
	Organization/ Process/ Policy		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	**O**
	Data		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	**O**
	HW		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	**O**

제4절 SW 신기술 안전 사고 조사 결과의 적용

1. SW 신기술 사고 유발 위험원 유형화와 대응

통상적으로 심도 있고 정확한 사고 원인 분석을 위해서는 SW 안전 전문가, 도메인 전문가, 사고 발생 기관 관계자 등이 조사팀을 구성하여 사고 원인 조사가 이루어져야 한다. 보통 이러한 조사는 사고당 짧게는 수개월에서 길게는 수년에 걸쳐 진행한다. 우선 관련 증거(evidence)를 수집하고, 관련 문서, 산출물들을 포괄적으로 취합 분류한다. 취합된 산출물에 대하여 기술적 분석을 진행하고, 사고 관계자들에 대한 시스템 역할별로(스폰서, 관리자, 개발자, 형상 관리자, 테스터, 사용자, 운영자 등) 심층 인터뷰가 이루어진다. 필요에 따라 사고 재현을 실시하고, SW 산출물 종류에 따른 심층 분석을 진행한다. 부족한 산출물에 대해서는 역공학을 통하여 재생하여야 하며, 변경 영향분석, 시뮬레이션 등 다양한 기술적 접근이 진행되어야 한다. 그리고 파악된 원인에 대하여 왜 사전에 HAZOP, FMEA, FTA와 같은 위험 분석이 이루어지지 않았는지 등에 대한 안전 공학적 분석도 필요하다.

하지만 본 연구에서는 주어진 증거산출물이 언론 기사만을 토대로 추론하여야 하므로 완전한 사고 원인 분석은 불가능하다. 다만 전문가로서의 추측만이 가능하다고 볼 수 있다. 본 연구는 그 목적이 발생한 사고들에 대한 원인 분석이 목적이 아니라 SW 신기술에 대한 향후 사고 예방을 위한 필요 조치, 정책 등을 파악하는데 그 목적을 가지므로, 이러한 전문가적 추론만으로도 그 가치가 크다고 볼 수 있다.

또한 제시한 분석 관점은 모두 각기 개별적이며, 상호 배제적이지 않으므로 다수의 잠재 원인이 도출될 수 있음을 밝힌다. 효과적인 전문가 추론을 위해 상기 방법론을 다음과 같은 질의 형식으로 전문가 의견을 취합했다.

[SW 신기술 사고 유발 위험원 유형(SW-ICAM)을 이용한 전문가 의견 조사 질의서]

질의 1. Safety Mechanism : Safety Mechanism 부재 혹은 오작동의 가능성이 있는가? 있다면 왜 가능성이 있다고 판단하는지 의견을 주시고 세부 요인를 선택해 주십시오. 그리고 어떠한 Safety Mechanism이 존재했어야 하는지 기술하시기 바랍니다. 또한 필요한 Safety Mechanism을 위해 사전에 어떤 잠재 위험 분석이 있었어야 했는지 설명하여 주시기 바랍니다.

질의 2. Actor 오동작 : 사용자 혹은 운영자의 오류, 연계시스템의 오류가 의심되는 경우 세부 요인를 선택하시고, 그 이유를 기술해 주십시오. 사람의 실수 혹은 미숙으로 인한 경우, 예측 가능한 실수(Reasonable Foreseeable Misuse)였는지 아니면 전혀 예측 불가능한 실수였는지 의견을 기술하여 주시기 바랍니다. 또한 악용에 대한 것인지에 대해서도 기술하여 주십시오.

질의 3. Environmental 요인 : 시스템의 작동 환경, 사용자/운영자의 환경 요인 혹은 기타 환경적 요인에 사고로 추정될 경우 해당 세부 요인를 선택하시고, 그 이유를 기술하여 주시기 바랍니다.

질의 4. Organizational/Process/Policy 요인 : 먼저 의심되는 세부 요인들을 모두 선택하시고 그 이유를 설명하여 주시기 바랍니다. 다수의 세부 요인가 선택 가능합니다.

질의 5. SW System 요인 : 기능 혹은 비기능적 요건에 대한 오류가 추정될 경우 세부 요인를 선택 바랍니다. 주어진 정보가 부족하여 세부 요인를 선택 못 할 수 있습니다. 이런 경우 그냥 기능, 혹은 비기능으로만 구분합니다.

질의 6. Data 요인 : 데이터 오류가 의심되는 경우 먼저 의심되는 세부 요인들을 모두 선택하시고 그 이유를 설명하여 주시기 바랍니다. 다수의 세부 요인가 선택 가능합니다.

질의 7. HW 요인: HW 오작동이 의심되는 경우 먼저 의심되는 세부 요인들을 선택하시고 그 이유를 설명하여 주시기 바랍니다. 주어진 정보에 따라 세부적으로 구분을 못 할 수 있습니다. 이러한 경우 HW 오류로만 언급하시기 바랍니다.

본 연구에서는 총 37건의 SW 신기술 사고 조사 보고서 각각에 대한 전문가 의견을 받아 최종적으로 SW 신기술 사고 유발 위험원 유형과 대응시켜 어떠한 부분이 SW 신기술 사고 완화 및 예방에 필요한 것인지를 도출했다. 세부 전문가 의견은 SW 신기술 관련 사고의 당사자에게는 민감한 요인이기 때문에 별도로 공개하지는 않으며, SW 신기술 사고 유발 위험원에 대응시킨 결과로 갈음한다. 37건의 SW 신기술 사고 조사 보고서는 [별첨]을 참고하면 된다.

2. SW 신기술 사고 유발 위험원 유형화와 대응 결과

<표 3-8> SW 신기술 사고 37건의 SW 신기술 사고 유발 위험원 유형 대응 결과

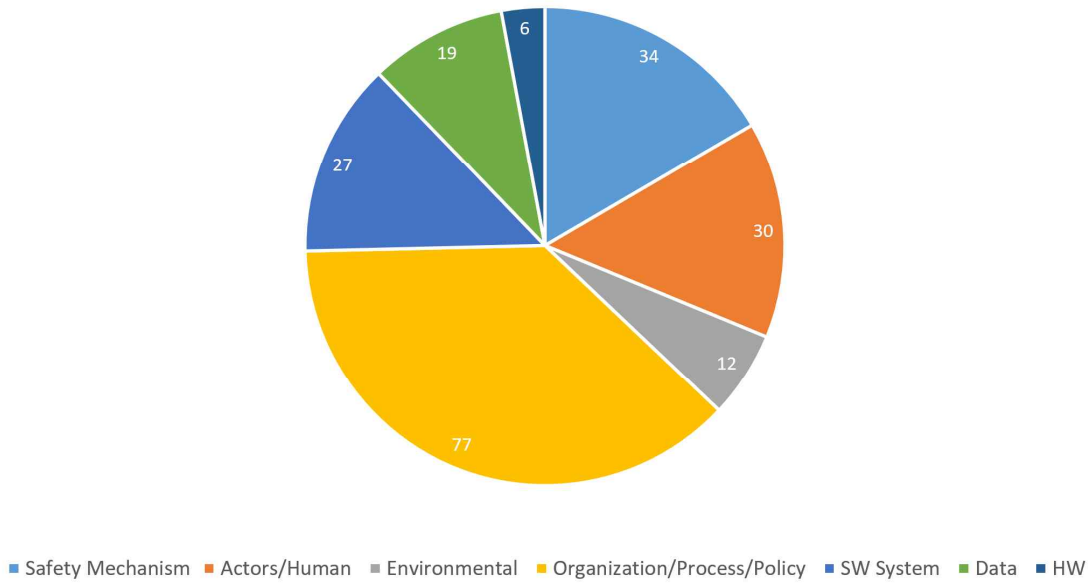
위험원 유형 \ 분 야			AI	메타버스	블록체인	합계
Safety Mechanism	Safety Mechanism 부재		12	8	7	**27**
	Safety Mechanism 오류	안전 기능 오류	4	1	0	**5**
		상호 작용 오류	1	0	0	**1**
		타이밍 오류	1	0	0	**1**
		Value 오류	0	0	0	**0**
Actors/ Human 오동작	실수		2	1	2	**5**
	미숙		1	3	1	**5**
	악의적 사용		7	5	8	**20**
	연계(Peer)시스템 오류		0	0	0	**0**
Environmental 요인	시스템 작동 환경		2	0	0	**2**
	사용자/운영자 작업 환경		1	3	0	**4**
	기타 환경 조건		2	4	0	**6**
Organization/ Process/ Policy 요인	사용자/운영자 교육 미흡 (악의적 사용이 아닌 경우)		2	2	0	**4**
	조직 구조상 문제 혹은 이슈		0	2	0	**2**
	의사소통 문제		0	0	0	**0**
	목표 충돌 문제		2	0	1	**3**
	사용/운영 절차 미흡		4	3	0	**7**
	유지 보수 미흡		1	0	0	**1**
	위험 관리 미흡		11	7	7	**25**
	변화 관리 미흡		1	1	0	**2**
	외주 관리 미흡		0	0	0	**0**
	조직 문화 이슈		0	0	0	**0**
	안전 제도 미흡		8	9	10	**27**
	악의적 사용 혹은 잠재 위험에 대한 교육체계 미흡		1	1	4	**6**
SW System 요인	기능 오류	사용자 기능 오류	7	0	2	**9**
		상호작용 오류	1	1	0	**2**
		타이밍 오류	0	1	0	**1**
		Value 오류	0	0	0	**0**
	비기능 오류	Platform 오류	0	0	0	**0**
		Security 오류	1	3	4	**8**
		Performance & Resource오류	0	0	0	**0**
		안전 요건 부재	1	3	3	**7**
Data 요인	데이터 부적합		5	3	1	**9**
	데이터 불충분		5	0	0	**5**
	데이터 편향		5	0	0	**5**
HW 요인	노화		1	0	0	**1**
	기준 부재		2	1	0	**3**
	기준 미달		1	1	0	**2**

총 37건의 SW 신기술 사고 조사 보고서에 대한 SW 신기술 사고 유발 위험원 유형을 대응한 결과는 <표 3-8>과 같다. 위험원 유형 대응은 SW 안전 전문가의 협의를 통해 진행되었으며, 하나의 사고 조사 보고서가 다양한 항목에 대응할 수 있다.

대응 결과 Safety Mechanism 부재(27건, 72.97%), 안전 제도 미흡(27건, 72.97%), 위험 관리 미흡(25건, 67.57%), 악의적 사용(20건, 54.05%)가 주요한 위험원 유형으로 드러났다. 이것이 시사하는 바는 SW 신기술에 대한 SW 안전의 역할이 사고의 완화와 예방에 매우 중요한 역할을 수행해야 한다는 것이다.

특히 악의적 사용의 경우 SW 신기술 안전 사고의 절반 이상을 차지하여 무분별한 남용을 제재할 수단이 필요할 것으로 분석된다. 이는 관련법의 제정으로 해결될 수 있는 부분이지만 선제적으로 막을 수단은 공급자의 노력이 필요한 부분이기 때문에, SW 신기술로 사업을 영위하는 주체의 윤리적인 인식 개선도 필요할 것으로 판단된다.

SW 신기술 사고 유발 위험원 유형의 7대 대분류 관점에서의 비중은 다음 [그림 3-7]과 같다.



[그림 3-7] SW 신기술 사고 유발 위험원 대분류의 비중

제4장 SW 신기술 안전 대응 동향

제1절 개 요

본 연구에서는 SW 신기술의 주요 분야인 AI, 블록체인, 메타버스 기술에 대한 사고의 유형과 그 원인에 대하여 모형화하였다. 이러한 신기술 분야는 다양한 지속적 기술 연구 개발과 함께 산업 경제 문화 사회 등의 분야에서 여러 서비스 형태로 실현되고 그 부가가치를 창출하려는 시점에 있다. 그러나 앞서 조사 분석에서 살펴본 바와 같이 다양한 형태의 안전 사고들이 발생하고 있으며, 이로 인한 안전 위협이 점점 증가하고 있는 추세이다.

이러한 안전 위협에 신속하게 대응하고, 보다 안전한 AI, 블록체인, 메타버스 기술이 사회 활동을 지원하게 될 때, 그 기술의 가치는 크게 증가할 것이다. 따라서 해당 분야에 대한 안전 대응 기법에 대한 심도 있는 연구 개발과 응용으로의 적용은 시급한 문제이다. 이러한 관점에서 AI, 메타버스, 블록체인에 대한 글로벌 안전 대응 동향을 주요 국가별 정책 혹은 표준화나 법 제도 등의 관점에서 살펴보기로 한다. 논의의 원활한 전개를 위하여 기술 자체에 대한 배경지식은 생략하겠다.

제2절 인공지능

1. 기술적 표준화 동향

ISO/IEC JTC 1, IEEE 등 다양한 단체에서 AI 기술 표준화에 앞장서고 있다. 이 단락에서는 각 단체의 표준화 현황과 각 단체의 표준화 대상을 중심으로 분석한다.

1) ISO

ISO는 공적 표준화 기구로, 참조 표준을 개발하여, 용어 정의, 요구사항 등을 정의하는 표준화 기구다. 2018년 4월 정보 기술 관련 ISO와 IEC의 합동 표준화 기구 ISO/IEC JTC 1에서는 AI 관련 표준화를 담당하는 SC 42 Artificial Intelligence(ISO/IEC, 2017)를 설립했다. 현재 35개 참가국과 15개 참관국을 포함해 50개국이 회원으로 참여하고 있

다. SC 42는 현재 5개의 WG (Working Group), 1개의 JWG (Joint Working Group), 1개의 AG (Advisor Group), 3개의 AHG (Ad-Hoc Group)을 포함하고 있다. 각 조직이 하는 역할은 다음 <표 4-1>과 같다.

SC 42는 현재까지 14개의 표준을 제정하고, 25개의 표준을 제정 중이다. WG 1는 출범 당시부터 존재하던 것으로, AI의 기반 표준을 담당하고 있다. 현재까지 AI 기술의 개념과 용어를 정리한 표준(ISO/IEC 22989:2022)과 머신러닝을 활용한 AI 시스템 프레임워크를 정리한 표준(ISO/IEC 23053:2022) 등 2개를 제정했다. WG 2는 JTC 1의 WG 중 하나였던 빅데이터 관련 WG를 이전하여 세워진 것이다. 빅데이터 개요(ISO/IEC 20546:2019), 참조 아키텍처(ISO/IEC TR 20547-1,2,3,5) 등 5가지의 표준을 제정했으며, 데이터 품질(ISO/IEC CD 5259-1,2,3,4,5), 라이프사이클(ISO/IEC DIS 8183) 표준 등이 제정중이다. WG 3은 AI의 신뢰성 표준을 담당한다. 신뢰성과 관련된 표준(ISO/IEC TR 24028:2020)은 물론, AI의 편향(ISO/IEC TR 24027:2020), 견고성 측정(ISO/IEC TR 24029-1:2021, ISO/IEC DIS 24029-2), 윤리적, 사회적 우려(ISO/IEC TR 24368:2022) 등 다양한 분야를 담당한다. WG 4는 유즈케이스와 다양한 어플리케이션을 관장한 표준을 담당한다. 현재 유즈케이스와 관련된 표준(ISO/IEC TR 24030:2021)이 제정되었으며, 빠른 개정 작업에 돌입했다. WG 5는 AI 시스템의 접근 방법과 관련된 표준을 담당하여, 현재 1개(ISO/IEC TR 24372:2021)의 표준을 제정했다.

다른 SC와의 협업으로 이루어지는 JWG는 현재까지 두 종류가 있었다. SC 40(IT 서비스 관리 및 IT 거버넌스)과 이루어진 JWG에서는 조직의 AI 활용이 거버넌스에 미치는 영향에 관한 표준(ISO/IEC 38507:2022)를 제시했으며, SC 7(SW, 시스템 공학)과 이루어지고 있는 JWG에서는 현재 AI 시스템을 위한 테스팅(ISO/IEC AWI TS 29119-11)과 관련된 표준이 제정 중이다.

이외에도 그래픽(SC 24), 미디어 코딩(SC 29), 데이터 관리(SC 32), 문서 설명 및 처리 언어(SC 34), SC 36(학습, 교육, 훈련을 위한 정보 기술), SC37(생체 인식), SC41(IoT와 디지털 트윈) 세부 그룹과 공동 작업하고 있으며, TC (Technical Committee) 중에서는 항공기, 의료 등 다양한 분야와 협업하고 있다.

이와 별개로 자율주행 등 AI을 사용하는 각 산업체 분야의 표준이 추가적으로 ISO에서 마련되었거나, 마련되고 있다. 다만, SC 42에서 중점적으로 표준화하는 분야는 AI 시스템으로, AI가 포함된 시스템만을 중심으로 표준화하고, 산업체별 표준은 각각 관리하는 것으로 보인다.

<표 4-1> SC42의 세부 조직

코드명	조직
WG 1	기반 표준 (Foundational Standards)
WG 2	데이터 (Data)
WG 3	신뢰성 (Trustworthiness)
WG 4	유즈케이스와 어플리케이션 (Use cases and applications)
WG 5	AI 시스템의 전산 방법론과 특성 (Computational approaches and computational characteristics of AI systems)
JWG 1 (종료)	SC 40(IT 서비스 관리 및 IT 거버넌스)과의 Joint Working Group AI의 거버넌스 영향 (Governance implications of AI)
JWG 2	SC 7(SW, 시스템 공학)과의 Joint Working Group AI 기반 시스템 테스팅 (Testing AI-based systems)
AG 1 (종료)	AI 관리 시스템 표준 (AI Management system standard)
AG 2 (종료)	AI 시스템 엔지니어링 (AI systems engineering)
AG 3	AI 표준화 로드맵 (AI standardization roadmapping)
AHG 1	보급과 확산 (Dissemination and outreach)
AHG 2	SC 38 (클라우드 컴퓨팅과 분산 플랫폼)과의 연락 (Liaison with SC 38)
AHG 3 (종료)	지능 시스템 엔지니어링 (Intelligent systems engineering)
AHG 4	SC 27(정보 보안, 사이버 보안과 사생활 보호)과의 연락 (Liaison with SC 38)
AHG 5 (AG3로 이관)	AI 표준화 동향 및 로드맵 (AI standardization landscape and roadmap)

2) IEEE

IEEE는 사실 표준화 기구로 실제 기술 개발과 밀접한 표준을 개발하는 기구다. 기존 자동화 시스템이나 적응 시스템의 표준들 중 AI으로 확장된 표준들을 비롯해 AI 및 자율 시스템의 표준들까지 64가지의 표준(IEEE SA, 2022)을 관리하고 있다. AI 모델의 정의, 로보틱스와 관련된 다양한 AI 시스템의 표준 등을 IEEE 2941을 비롯한 다양한 표준에서 관리하고 있다.

현재 IEEE에서 제일 집중하고 있는 것은 윤리적 문제를 해결하기 위한 모델 프로세스에 관한 P7000-2021 표준을 비롯하여 7001~7015번 표준들로, AI 시스템의 윤리적인 문제들과 투명성, 프라이버시, 편향, 보안, 안전성 등 가치 기반 공학(Vienna, 2021)에 집중하고 있다. 가치 기반 공학은 IT 시스템을 인간 중심적으로 변환하고, 이해관계자들의 가치를 더 잘 반영하도록 하는 것으로, 특히 7000번 표준에서 강조하고 있는 부분이다. 해당 표준에서는 IEEE의 AI 시스템을 대상으로 하는 윤리 인증 프로그램(Ethics Certification Program for Autonomous and Intelligent Systems Certification, ECPAIS Certification)의 기준을 제공하고, 윤리적인 문제를 프로그램과 시스템 설계에 엄격하게 우선할 것을 제시하고 있다. 특히 가치 중심적 방법론을 기반으로 피해 계산에 잠재적인 가치 피해도 고려하도록 추진하고 있다.

2. 글로벌 법제화 동향

AI 관련 법제화는 대부분 위험성을 관리하고 규제하기 위한 필요에서 탄생했다. 해당 필요는 산업체와 사회 각 층에서 제안되었으나, 법률로 제정된 곳은 아직 유럽연합이 유일하다. 다른 국가에서는 지침과 보고서 수준에 그치고 있다. 관련 지침들에서 규제하고 있는 분야는 크게 안전, 개인정보와 관련된 보안, AI 윤리와 관련된 분야가 있다. 이 단락에서는 AI법을 제정한 유럽연합을 비롯해 국가별 법제화 관련 동향을 소개한다.

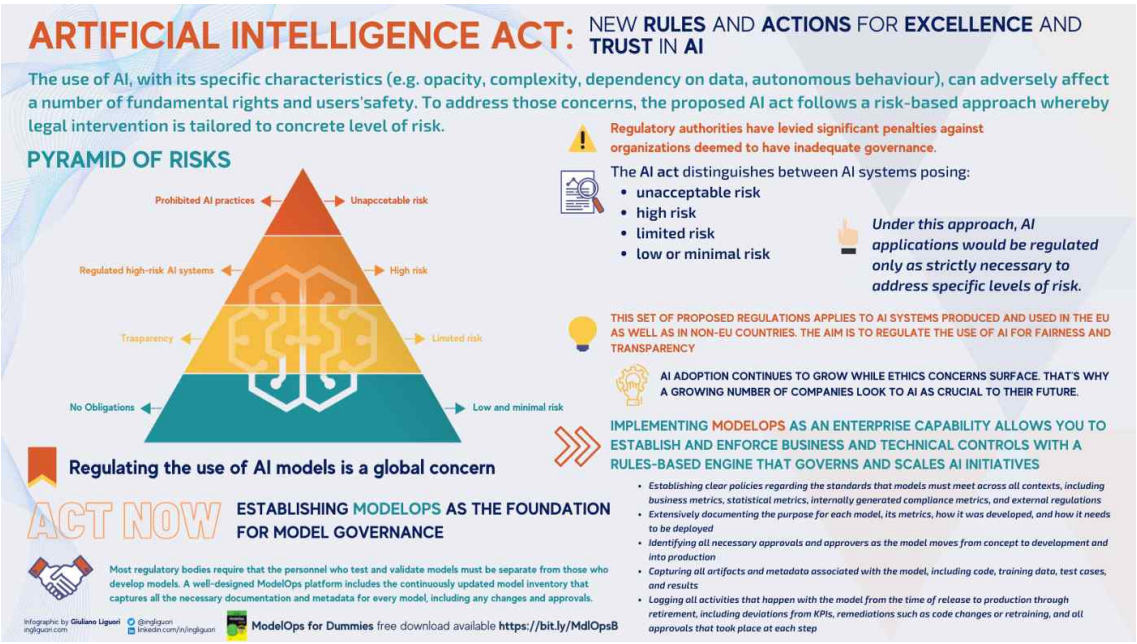
1) UNESCO

2021년 UNESCO 41차 총회에서 AI 윤리에 관한 권고(The Recommendation on the Ethics of Artificial Intelligence)(UNESCO, 2022)가 통과되었다. 해당 권고는 전문과 8개 장, 141개조로 구성되어 있다. 3장과 4장이 해당 권고의 핵심 부분으로, 3장에서 추구하는 핵심 가치와 원칙들을 명시하고, 4장에서 각 분야별 필요한 정책을 나열하였다. 이 권고에서 지향하는 핵심 가치는 인권 및 인간 존엄성, 환경과 생태계, 다양성과 포용성, 평화롭고 공정한 사회의 네 가지다. AI 시스템으로 물리적, 경제적, 사회적, 정치적, 문화적, 정신적으로 인간이나 공동체가 해를 입거나 종속되어서는 안되며(제14조), AI 시스템과의 상호작용으로 인권이나 기본적 자유가 침해되거나 남용되어서는 안된다는 내용(제15조) 등을 담고 있다. 이러한 네 가지 가치를 위하여 비례성, 안전성과 보안, 공정성, 지속가능성, 개인 정보 보호, 인간의 감독, 투명성 및 설명 가능성, 책임

과 의무, 대중의 인식, 다중 이해 관계자 체제 등 열 가지 원칙을 통해 구체화하고 있다. 4장에서는 윤리적 영향 평가, 데이터 정책, 환경 및 생태계, 젠더, 문화, 교육, 경제, 복지 등 11개 분야에서 핵심 가치를 지키기 위한 정책을 구체적으로 제시하고 있다. 특기할 것은, 인간 중심적 AI 시스템을 강조했다는 점이다. AI 시스템에 기반한 것이 아닌, 인류 전체의 본질적인 핵심 가치에서 출발해 AI 시스템으로의 정책으로 확장되어 나가는 것이 특징이다.

2) EU

2022년 기준 유럽연합은 AI법(Artificial Intelligence Act)을 제정했다. AI법은 [그림 4-1]과 같이 4가지 위험 수준에 따라 규제를 다르게 하고 있다(Liguori, 2022).



[그림 4-1] AI ACT 설명

예측할 수 없는 위험(Unpredictable risk), 높은 위험(High risk), 제한된 위험(Limited risk), 낮은 위험(Low and minimal risk)로 나누어지고, 그 수준에 따라 테스트 데이터 투명성, 사이버 보안 요구, 인가의 감독 등의 의무가 다르게 적용된다. 예를 들어 챗봇과 같이 인간과 상호작용하는 시스템이나, 생체 데이터를 기반으로 감정 등을 추정하는 시스템, 딥페이크 등의 시스템은 기계와 인간이 상호작용하고 있음을 인식할 수 있는 최소한의 투명성 의무가 필요하다. 교육, 고용, 공공 서비스, 의료 등에 사용되는

AI는 높은 위험으로 평가되어 출시 이전에 데이터셋 검증과 평가를 엄격하게 받고, 당국에 상세한 문서를 제공해야 하며, 인간의 자유 의지를 회피하여 행동을 조작하는 AI 시스템, 정부가 인간을 평가하는 사회 평가 시스템 등은 예측할 수 없는 위험에 해당되어 사용 및 배포가 전면 금지된다. 이러한 법안의 규칙을 불이행하는 경우, 최대 글로벌 총 매출액의 6%에 해당하는 금액을 벌금으로 부과받을 수 있다.

제3절 메타버스

1. 기술적 표준화 동향

메타버스 표준은 아직 제정되고 있지 못한 상태이다. 메타버스와 관련된 다양한 기술 분야, 즉 자연어처리, 블록체인, 비전 및 영상처리, 네트워크, 디지털 트윈과 같은 분야에서의 표준들은 존재하지만, 이들이 SW 안전에 대한 고려를 하지는 못하고 있다.

메타버스 표준 포럼(MetaverseSF, 2022)은 2022년 6월 21일 결성되어, 메타버스 관련 개방형 표준을 개발할 목적으로 크로노스 그룹(Khronos Group)에 의해 주도하고 있다. 아직은 시작 단계이며, World Wide Web Consortium, Open Geospatial Consortium, OpenAR Cloud, Spatial Web Foundation, OASIS Consortium 등의 연합과 함께 현재 650명의 멤버가 참여하여 표준화를 위한 Working Group를 구성하고 있다. 크로노스그룹은 2000년 ATI 테크놀로지스, Discreet, Evans & Sutherland, 인텔, 엔비디아, 실리콘 그래픽스, 썬 마이크로시스템즈의 주도로 설립되었으며, 현재 약 120개 기관이 참여하고 있다. 미국 오리건주에 본사를 둔 비영리 산업체 컨소시엄으로 로열티가 없는 개방형 표준 개발을 하고 있다.

이 포럼의 주요 관심사는 웹과 몰입도를 갖는 공간 컴퓨팅을 갖는 복잡한 연결을 메타버스의 특성으로 규정하고, 메타버스 상호 운영성에 관련된 표준을 일차적으로 개발하려는 목적을 갖는다. 국내 기업중에서는 LG, 네이버, 한국전파진흥협회, 맥스트(MAXST), 맥슨(MAXON) 등이 참여자 목록에 올라 있다.

2. 글로벌 법제화 동향

1) 미국

미 하원 보고서 'Investigation of Competition in Digital Markets' (2020)(Nadler , 2020)은 다음과 같은 플랫폼 규제 방향을 제시하고 있다.

- 거대 플랫폼의 구조 조정(기업 분할 등) 및 인접 시장 진출 제한
- 자기 우대 검색 제한, 동일 서비스에는 동일 계약 조건, 결합 판매 금지 등
- 상호 운용성 및 데이터 이동성 등 개방적 서비스 접근 촉구
- 판매자 지배력: 시장 점유 30%, 수요자 지배력: 25%로 양적 기준 정의 등

하원 보고서의 후속 조치로, 2021년 6월 11일 미 하원이 플랫폼 규제 5개 법안(플랫폼 규제 패키지 법안)을 발의했다. 플랫폼 규제 대상 기준은 ① 미국 기반의 활성 이용자 월 5천만명, 또는 활성 사업 이용자 월 10만명, ② 시가총액 6000억 달러 이상, ③ 온라인 플랫폼에서 재화·용역 판매를 위한 중요한 거래 상대방(사실상 GAFA: Google, Apple, Facebook, Amazon)을 규제한다. 또한 플랫폼 제공자의 자사 재화·용역 판매 중단, 인수합병의 경쟁 제한성 없음에 대한 입증책임 부과, 자사제품 우대, 소셜 미디어 데이터 이동성 보장 등을 주요 내용으로 한다.

플랫폼 규제는 모두 메타버스 플랫폼의 발전 경로에 영향을 미칠 수 있지만, 특히 '플랫폼 위의 플랫폼'을 통제할 수 있는 OS/디바이스 플랫폼에 대한 규제가 결정적인 영향을 미칠 수 있다는 측면에서, Open App Market Act 시행 여부에 주목할 필요가 있다(최계영, 2021).

OS/디바이스/앱스토어가 결합된 게이트키퍼는 서비스 플랫폼에 대한 사실상의 인허가, 제공되는 기술/기능적 특징, 커미션 수준 등을 좌우함으로써 디지털 콘텐츠 제공 플랫폼으로서의 메타버스에 대한 최종적 권한(ultimate authority)를 보유하게 되는 문제가 있다. 또한 iOS와 안드로이드 플랫폼 간 전환비용이 높고, UI/편의성 등의 차이로 인해 웹 또는 제3의 앱스토어를 통한 우회가 현실적으로 어려워 개발자들이 OS 게이트키퍼의 영향력에서 벗어나기 어렵다.

개방 앱마켓 법안(Open App Market Act)의 주요 내용은 앱 경제에서의 게이트키퍼의 시장 지배력을 제한하기 위함을 그 목적으로 하며, 미국내 5천만명 이상 이용하는 앱스토어 운영자를 규제 대상으로 한다.

2) 유럽

유럽은 메타버스를 위한 포괄적인 법안으로 DMA(Digital Market Act)와 DSA (Digital Service Act)를 적용하고 있다. DMA 5조는 게이트키퍼가 반드시 준수해야만 하는 의무를 열거<표 4-2>하고 있으며, DMA 6조는 추후 개선이 가능한(susceptible of being further specified) 사안들로, 게이트키퍼와의 협의 대상을 정의하고 있다(최계영, 2021).

<표 4-2> DMA 5조 : 게이트키퍼가 준수해야 하는 의무

주요 내용	해당 게이트 키퍼(예상)
(a) 게이트 키퍼의 다른 서비스나 써드 파티로부터의 개인 데이타를 결합하는 행위 금지하고 최종 이용자가 케이트키퍼 다른 서비스에 가입하도록 요구 금지	Facebook, Google
(b) 사업 이용자의 써드파티 중개 이용을 허용	Amazon
(c) 사업 이용자다 게이트키퍼 통하지 않고 판매 촉진할 수 있어야 하고 이에 대한 최종 이용자의 접근 및 이용이 가능해야 함.	Apple
(d) 공공기관을 통한 이의제기 가능	All Gate keepers
(e) 사업이용자에 게이트키퍼의 식별 서비스 이용 금지를 요구	Facebook, Google
(f) 사업 & 최종 이용자에 게이트 키퍼의 다른 핵심 서비스 이용을 요구하는 행위 금지	Facebook, Google
(g) 광고주 및 광고 퍼블리셔에 가격 등 요구 정보를 제공할 의무	Facebook, Google

출처 : EC, 2020에서 재구성

GDPR(General Data Protection Regulation) 개정 추진 : GDPR은 개인 정보를 규제하지만 메타버스를 효과적으로 규제하려면 개정이 필요하다. 예를 들어 데이터 유출로 인해 암호 화폐가 손실되는 경우 메타버스 소유자와 암호화 플랫폼과 같은 제3자 서비스 제공업체에 더 많은 책임을 부여하여 사용자가 안전하게 거래할 수 있도록 하고자 한다. 또한 DSA(Digital Services Act)은 데이터 사용에 대한 제한 사항을 명시적으로 언급하도록 규정하고 있다. 아바타 및 관련 콘텐츠의 신원을 규제하는 데 중요한 역할을 수행한다.

3) 일본

일본에서는 일본의 특정 디지털 플랫폼의 투명성 및 공정성 향상에 관한 법률을 2020년 6월 3일 공포하고, 2021년 2월 1부터 시행(METI, 2021)하고 있다.

일본 플랫폼 공정화법의 주요 내용은 다음과 같다.

기본 방침	- 이용자 편익 증진에 기여하며 일본 경제 사회 활력 제고와 지속적 성장 발전 - 디지털 플랫폼 사업자가 투명성·공정성 향상을 위해 자발적·적극적으로 노력 - 디지털 플랫폼의 독창성을 충분히 발휘할 수 있도록 사업자·중개자·이용자 간 거래 관계에 있어 상호이해 촉진 도모
규제 대상	- 규제대상: 디지털 플랫폼 제공자 (제2조, 제5호, 제4조) - 상품판매 온라인 사업자(일본 내 매출 3000억엔 이상)과 앱 스토어 사업자(2000억앤 이상)가 규제 대상 - 대표적인 일본 기업은 라쿠텐, 야후재팬, 아마존 재팬 등이며 해외 기업은 구글, 애플, 페이스북 등의 거대 기업을 겨냥함
규제 내용	- 거래 조건 등 정보 공개 : 이용자에 대한 계약 조건 공개 및 변경 시 사전 통지 의무화 : 소비자 구매 데이터 등을 어떻게 취득·이용할지도 공개 : 보안상 이유로 공개할 수 없는 경우, 적절한 예외 조항 마련 - 절차·체제의 자발적 정비 : 정보 공개가 이루어지지 않는 경우에는 권고 등 행정 조치 : 규제 대상 사업자는 경제산업성 장관이 정하는 지침에 입각하여 절차·체제 (사업자 문의에 대한 매뉴얼 정리 등) 정비 실시 - 운영 상황 보고 및 모니터링 : 거래 조건 등 정보 공개, 절차·체제의 자발적 정비에 대한 자체 평가 보고서를 경제산업성에 매년 제출 의무 : 경제산업성은 보고서를 분석·평가, 관계자 간 유관 정보 공유, 상호 이해 촉구 및 모범 사례 제시 등
기타 규칙	- 공정거래위원회와 연계해 부적절한 행위나 위반 시 적절한 조치 - 거래 규칙을 총괄하는 전문 담당자 배치 - 일본 기업뿐만 아니라 해외 플랫폼 사업자에게도 동일하게 적용 - 시행 3년 후 점검·검토를 통해 필요한 후속 조치 법안 마련

자료 : 일본 경제산업성(經濟産業省), 2020.1.26.

4) 메타버스에 적용하는 현행법 (Olga, 2022)

현재 메타버스 또는 기타 가상세계를 구체적으로 규제하는 법률은 없다. 그러나 메타버스에도 저작권법, 명예 훼손법, 계약법 등 인터넷에 적용되는 것과 동일한 일반법이 적용된다.

- 저작권법(Copyright Law): 메타버스에서 저작권법은 일반적으로 아바타, 가상 건물 및 디지털 아트웍과 같은 사용자 생성 콘텐츠에 적용되지만 메타버스에서 실제 세계에서 저작권이 있는 저작물과 같은 것을 만드는 경우에도 저작권을 침해할 수 있다.
- 지적 재산권법(Intellectual Property Law): 지적 재산권법은 발명, 상표 및 기타 마음의 창조물을 보호하는 것으로, 메타버스에서 지적 재산권법은 아바타, 가상

건물 및 디지털 아트워크와 같은 가상 상품 및 서비스에 적용될 수 있다.

- 계약법(Contract Law): 계약법은 계약의 형성과 이행을 규율한다. 메타버스에서는 이용자 간의 계약에 대해 계약법을 적용하며, 이러한 계약에는 가상 상품 판매에서 가상 자산 임대에 이르기까지 모든 것이 포함될 수 있다.
- 불법 행위법(Tort Law): 불법 행위법은 개인 상해 및 재산 피해와 같은 민사상의 잘못을 다루며, 메타버스에서 불법행위법은 다른 사용자에 의해 발생한 모든 피해에 적용된다. 여기에는 신체적 상해, 정서적 고통 및 재산 피해가 포함될 수 있다.
- 명예 훼손법(Defamation Law): 명예 훼손법은 사람들에 대한 거짓 및 해로운 진술로부터 사람들을 보호한다. 메타버스에서는 타인 또는 회사를 비방하는 사용자 제작 콘텐츠에 명예훼손법이 적용된다.

제4절 블록체인

1. 주요국의 정책 동향

1) 유럽의 블록체인 정책

① 블록체인 정책과 데이터 보호 규정에 대한 유럽 의회의 정책 (European Parliament Report: Blockchain and the General Data Protection Regulation) (EUP, 2019)

유럽의 입장은 블록체인은 일종의 기술이라는 점에 초점을 맞추는 것이다. 실제로 이 기술의 하나의 버전만 존재하는 것이 아니라 다양하게 변화된 버전들이 존재한다. 이것은 또한 아래 분석에서 충분히 강조될 바와 같이 분산 원장과 유럽의 GDPR(General Data Protection Regulation) 간의 호환성은 관련 블록체인 사용 사례의 특정 기술 설계 및 거버넌스 설정을 설명하는 상세한 사례별 분석을 통해서만 평가될 수 있음을 의미한다.

결론적으로 말하자면, 블록체인이 모두 유럽 데이터 보호법과 호환되거나 호환되지 않는다는 일반화 된 방식으로 결론 지을 수 없다는 것을 알 수 있었다. 데이터 보호법과의 호환성 문제는 case-by-case의 이슈라는 것이다.

단편적인 호환성 분석보다, 블록체인 기술은 각 케이스 마다 유럽 데이터 보호법과 호환성에 도달하기 위해 블록체인의 어떤 세부적인 기술이나 세팅을 활용할 수 있을지 검토하고 추진하는 것이 합당하다. 즉, 공개 혹은 사용 허가가 필요 없는(permissionless) 네트워크보다 EU 데이터 보호법과 호환되는 방식으로 private 하고 사용 승인된 블록체인을 설계하는 것이 더 쉽고 바람직하다.

이는 사용 허가된(permissioned) 네트워크의 참여자가 다른 사람에게 알려져 있기 때문에 적절한 책임 배분을 가능하게 하는 계약 관계와 같은 정의가 가능하기 때문이다. 더욱이, 이들 네트워크들은 public 및 permissionless 네트워크들과는 달리, 데이터를 합법(compliant)하게 취급하는 것과 같이 네트워크를 통한 제어를 가능하게 하는 방식으로 설계된다. 또한 어떤 행위자가 관련된 개인 데이터에 액세스할 수 있는지에 대한 제어가 가능하다. 이러한 보호 메커니즘은 공개 및 허가되지 않은 블록체인의 경우에는 해당되지 않는다.

유럽의 정책은 이러한 블록체인의 특성을 고려하여 별도의 블록체인을 위한 법률 제정을 하는 접근이 아니고, 기존의 관련 법률의 충족요건에 부합하도록 기술적으로 접근하는 방법을 취하고 있다. 자동차나 항공 분야와 같이 안전에 관련된 별도의 새로운 법이나 표준을 제정하고 있지는 않다.

이 연구에서는 블록체인 기술과 데이터 보호법의 전반적인 구조 사이에 상당한 이슈가 있음을 인지하고 있다. 또한 기술과 법적 틀 사이의 관계는 일반적인 방식으로 결정될 수 없으며 오히려 사례(use-case)별로 결정되어야 한다는 점을 강조한다. 반면에 이러한 기술은 GDPR의 목표 중 일부를 달성하는 데 도움이 될 수 있는 뚜렷한 이점을 제공할 수 있다고 보고 있다. 앞의 분석에 기초하여 다음과 같은 구체적인 정책 권고 제시하고 있다.

- 정책 옵션 1 – 규제 지침(regulatory guidance)

현재 유럽 데이터 보호법의 다양한 요소가 블록체인 맥락에서 어떻게 적용되어야 하는지에 대한 법적 확실성이 부족하다는 것이다. 이것은 두 가지 중요한 요인 때문이라고 말하고 있다. 첫째, 블록체인 기술의 기술적 구조와 거버넌스 배열이 GDPR의 요구사항과 대조되는 경우가 종종 있다는 것이다. 둘째, 규제를 블록체인 기술에 매핑 하려는 시도는 이 법적 프레임워크의 해석 측면에서 모호한 부분이 많다는 것이다.

실제로 GDPR이 구속력을 갖게 된 지 거의 일 년 후, 법적 제도는 주로 이전의 1995

년 데이터 보호 지침에 기반을 두고 있지만, 많은 핵심 개념이 아직 불분명하다는 것이다. 예를 들어, 익명화 또는 (조인트)데이터 컨트롤러의 개념과 같은 핵심 개념은 여전히 불안정한 상태로 남아 있다. 이러한 개념의 해석은 유럽연합의 다양한 감독 당국 간의 해석에 대한 합의의 부족으로 인해 쉽지 않다.

또한 블록체인 기술이 데이터 최소화 및 목적 제한과 같은 유럽 데이터 보호법의 핵심 가정 측면에서 쉽지 않은 측면이 있다는 것이다(그러나 동시에 이러한 원칙은 빅데이터 분석과 같은 현대 데이터 경제의 분야에서도 동일한 문제를 가지고 있다). 그럼에도 불구하고 이 연구에서 연구진들은 GDPR을 개정할 필요가 없다고 권고하고 있다. 이 규정은 기술적으로 중립적이며 빠르게 변화하는 데이터 경제에서 원론적인 원칙 기반 규제의 표현이기 때문이다. 원론적이고 어느 정도는 추상적이어야지 빠른 변화에도 모두 적용될 수 있다고 보고 있다. 따라서 블록체인 기술을 사용하고자 하는 사람들에게 법적인 부합성을 높이기 위해 필요한 것은 GDPR을 고치는 것보다는, 이러한 메커니즘이 사용되는 곳에서 특정 개념을 어떻게 적용해야 하는지에 대한 규제 적용을 위한 지침이다.

규제 지침을 통해 GDPR 적용 시 법적 부합성을 더 많이 제공할 수 있다고 보고 있다. 이것은 다양한 규제 이니셔티브의 형태를 취할 수 있다. 물론 필요한 경우 지나치게 추상적인 GDPR의 특정 부분을 좀 더 구체화하는 것도 필요할 수 있다. 이러한 규제 지침을 통하여 보다 더 투명한 법적 부합성을 확보할 수 있다고 보고 있다. 물론 다양한 사례에 대한 규제 지침을 개발하는 것이 쉬운 일은 아니다.

- 정책 옵션 2 - 행동 강령(code of conduct) 및 인증 메커니즘(certification mechanism) 지원

기술적으로 중립적인 법적 프레임워크로서 GDPR은 모든 기술에 적용할 수 있는 방식으로 설계되었다.

그러나 GDPR의 기술 중립성은 위의 분석에서 알 수 있듯이 개인 데이터 처리의 특정 사례에 적용을 어렵게 할 수 있다. 이는 중립성이냐, 적용의 용이성이냐의 문제이다. 이를 해결하기 위해서 인증 메커니즘을 활용할 수 있다. 실제로 인증 메커니즘과 행동 강령은 GDPR에서 특별히 언급한 구체적인 도구로서, GDPR의 가장 중요한 원칙을 개인 데이터가 처리되는 구체적인 상황에 적용하는 데 도움을 줄 수 있다. 일반적인 프로세스 표준 혹은 품질 표준을 해석할 때, 해당 심사 방법론을 보면 더 쉽게 해석할 수 있는 것과 같은 맥락이다.

- 정책 옵션 3 - 연구 자금 지원

규제 지침뿐만 아니라 행동 강령 및 인증 메커니즘은 GDPR의 특정 조항이 블록체인 기술과 관련하여 어떻게 적용되어야 하는지에 대한 많은 법적 확실성을 추가 할 수 있다.

그러나 이것이 GDPR에 대한 특정 분산 원장 사용 사례를 준수하기에 항상 충분하지는 않다. 따라서 다양한 연구를 통해 기술 및 거버넌스 구제책을 고안하고 설계에 따라 준수할 수 있는 블록체인 프로토콜을 사용한 실험 등 지속적인 연구가 뒷받침되어야 한다.

2) 미국의 블록체인 정책

① 암호화 화폐에 대한 미국의 기본 입장

2022년 미국 로이터 통신(Reuter, 2022)에 의하면 미국은 새로운 법 없이도 암호화 화폐를 규제할 수 있다는 전문가 의견이 나왔다. 미국의 규제 기관들은 현존하는 법을 이용하여 암호화 화폐와 같은 디지털 자산에 대한 규제가 가능하다는 입장을 취하고 있다. 미국 증권거래위원회인 SEC(Securities and Exchange Commission)와 CFTC(Commodity Future Trading Commission)는 CAP(Center for America Progress)에서 발행된 상기 내용의 보고서를 인용하며 암호화 화폐에 대한 지배(governance)를 하겠다는 입장이다.

아직까지 규제 기관들에서는 공식적으로 어떠한 방법으로 암호화 화폐, 특히 소위 “스테이블코인”이라고 불리는 화폐를 관리/규제할지 공표하고 있지는 않다. 현재 미국 U.S. Treasury Department는 이 이슈에 대한 해결을 2021년에 의회로 전가한 상황이다.

② 미국이 바라보는 Blockchain의 특성

미국에서 바라보는 블록체인 기술은 유럽과 유사하지만 실용적인 측면에서 좀 더 실무적인 관점이 많다고 볼 수 있다. 다음은 미국에서 바라보는 블록체인에 대한 주요 특성이다.

• Decentralized Blockchain

암호화 화폐와 기타 디지털 자산의 생성에 활용되는 주요 기술로서 블록체인의 Decentralized 기술과 Permissionless 기술을 볼 수 있다. 분산된 퍼블릭 블록체인에서는 누구든지 자금의 송금과 수신을 위해 소프트웨어를 다운로드 받을 수 있다. 특히 은행과 같은 중앙관리 조직이 없이 가능하다는 것이 특징이다.

• Centralized Blockchain

Centralized 되고 permission된 블록체인 기술은 기업이나 정부에서 고속의 기록관리를 위해 사용된다. IBM의 Hyperledger Fabric 소프트웨어를 사용한 블록체인 개발이 이에 속한다. 중앙화된 블록체인에 대한 접근은 제한되고, 권한이 있는 사람만이 접근할 수 있다, 이런 중앙화된 블록체인은 통상 몇 개의 컴퓨터 노드에 작게 분산되어 관리된다. 분산된 블록체인 대비, 중앙화 블록체인의 장점은 기존보다 모든 기록에 대한 처리를 빠르고 에너지 효율적으로 할 수 있다는 것이다.

③ 블록체인 산업에 대한 미국의 기본 입장

이와 같이 분산 혹은 중앙화된 블록체인에 대한 활용관점에서 매우 유연하게 다양한 업무 도메인에서 시도 적용되고 있다. 전체적으로 유럽보다 더 적극적으로 블록체인 발전을 가속화하는 입장으로 이해되며, 기본의 관련 영역 법을 통하여 관리하겠다는 입장이다.

2. 글로벌 법제화 동향

1) 유럽연합의 블록체인 관련 법제도 현황

① 유럽연합의 일반 데이터 보호 규정

유럽연합의 일반 데이터 보호 규정(GDPR)은 2018년 5월에 구속력을 갖게 된 데이터 보호법이다. 이 법은 기존의 1995년 데이터 보호 지침을 기반으로 한다. GDPR의 목표는 기본적으로 두 가지이다. 첫째, EU의 다양한 회원국 간의 개인 데이터의 자유로운 이동을 촉진한다. 둘째, 기본권 헌법 제8조의 정보 보호권을 바탕으로 기본권 보호 프레임워크를 확립하고 있다.

이 법적 프레임워크는 데이터 처리의 수단과 목적을 결정하는 주체인 데이터 컨트롤러에 대한 여러 가지 의무를 정립하고 있다. 또한 데이터 주체(개인 데이터와 관련된

자연인)에게 데이터 컨트롤러를 통해 시행할 수 있는 여러 가지 권리를 할당하고 있다.

② 블록체인과 GDPR 사이의 대립

유럽에서는 최근 몇 년 동안 블록체인 기술과 GDPR 간의 여러 대립이 있었다고 언급하고 있다. 세부적인 대립 현상은 아래에서 설명하겠지만, 기본적으로 우리가 바라보아야 할 관점은, 신기술이 나왔을 때 안전과 관련된 관점에서 현존하고 있는 규정이나 법 제도에서 바로 검증 액션들이 이루어지고 있다는 것이다. 이는 국내와는 좀 대조적이라고 볼 수 있는데, 국내에서는 신기술 자체에 환호하며 신기술이 만들 혁신에 더 열광하는 현상이 많다. 관련 규제를 언급하는 경우 신산업의 성장을 규제한다는 시각으로 비평하는 경우가 많다. 그럼 유럽의 경우를 살펴보도록 한다. 앞에서 언급한 블록체인과 GDPR 간의 대립은 두 가지 중요한 요인 때문이라고 볼 수 있다.

첫째, GDPR은 각 개인 데이터 포인트와 관련하여 데이터 주체가 EU 데이터 보호법에 따라 권리를 행사하기 위해 처리할 수 있는 적어도 한 명의 자연인 또는 법인 형식의 데이터 컨트롤러가 있다는 기본 가정을 기반으로 하고 있다.

그러나 블록체인은 종종 단일 액터를 많은 다른 플레이어로 대체하는 데 있어 분권화를 추진한다. 이것은 책임의 할당과 책임자에 대한 명확성을 어렵게 만든다. 특히 규정에 공동 통제권의 개념의 자체의 불확실한 실체 측면에서 그러하다. 이와 관련하여 또 다른 복잡한 요소는 최근의 판례법 발전에 비추어 볼 때, 어떤 주체가 공동 통제권의 자격이 있다고 정의하는 것이 법적 불확실성이 많다는 것이다.

둘째, GDPR은 GDPR 제16조 및 제17조와 같은 법적 요구사항을 준수하기 위해 필요한 경우 데이터를 수정하거나 삭제할 수 있다는 가정을 기반으로 하고 있다. 그러나 블록체인은 데이터 무결성을 보장하고 네트워크에 대한 신뢰를 높이기 위해 의도적으로 이러한 데이터 수정을 어렵게 하는 접근을 취하고 있다. 이는 데이터 보호에 법적인 불확실성이, 특히 EU 데이터 보호법 측면에서 매우 높아진다는 것이다. 예를 들어, GDPR 제 17조의 '삭제' 개념이 블록체인에서는 어떻게 해석되어야 하는지는 현재 불분명하다.

이러한 문제들은 여러 분야에서 일어나고 있다. 예를 들어, 공개키 및 트랜잭션 데이터와 같은 분산 원장에 일반적으로 저장된 데이터가 GDPR의 목적을 위해 개인 데이터로 자격이 있는지 여부에 대한 논쟁이 있다. 특히 문제는 암호화되거나 해시된 개인 데이터가 여전히 개인 데이터로 자격이 있는지에 대한 여부이다.

블록체인과 GDPR 간 논쟁의 또 다른 예는 데이터 최소화 및 목적 제한의 중요한 원칙과 관련이 있다. GDPR은 처리되는 개인 데이터를 최소한으로 유지하고 사전에 지정된 목적으로만 처리하도록 요구하지만, 이러한 원칙은 블록체인 기술에 적용하기가 어려울 수 있다. 분산 원장은 새로운 데이터가 추가됨에 따라 지속적으로 증가하는 추가 전용 데이터베이스이기 때문이다. 또한 이러한 데이터는 여러 다른 컴퓨터에 복제된다. 두 측면 모두 데이터 최소화 원칙의 관점에서 문제가 있다. 또한 개인 데이터 처리의 '목적'이 블록체인상에서 어떻게 반영되어야 하는지, 특히 이것이 초기 거래만 포함하는지 또는 체인에 배치된 후 개인 데이터의 지속적인 처리도 포함하는지 여부는 분명하지 않다.

또 다른 이슈는 삭제할 권리('잊혀질 권리')와 최근 몇 년 동안 가장 많이 논의된 블록체인 사이의 문제이다. 실제로 블록체인은 일반적으로 데이터의 (일방적)수정을 어렵거나 불가능하게 만들기 위해 의도적으로 설계되어 있다. 물론 이는 특정 상황에서 개인 데이터를 수정(GDPR 제16조에 의거)하고 삭제(GDPR 제17조에 따라)해야 한다는 GDPR의 요구사항에 온전히 부합한다고 보기 어렵다.

GDPR과 블록체인 사이의 문제에 대한 궁극적인 결론은 다음과 같다. 첫째, 블록체인 사용 사례의 기술적 특수성과 거버넌스 설계는 GDPR에 부합하기에 어려울 수 있다. 따라서 블록체인 설계자는 처음부터 이것을 인식하고 유럽 데이터 보호법을 준수할 수 있는 방식으로 각각의 사용 사례를 설계해야 한다고 주장하고 있다. 둘째, 그러나 블록체인이 규정을 준수하는 방식으로 어떻게 설계될 수 있는지에 대한 법적 확실성의 문제이다. GDPR이라는 렌즈를 통하여 블록체인을 바라보면, GDPR를 준수하기 위한 기술적으로 명확한 해답 찾기가 쉽지 않다는 것이다.

이 논쟁의 대부분은 블록체인과 유럽 데이터 보호법 사이의 긴장에 초점을 맞추었지만, 블록체인의 관점에서 GDPR 요건을 바라보면 해결책을 찾는 것이 더 용이해질 수 있다고 판단하고 있다.

③ 블록체인에 대한 유럽 집행위원회(European Commission)의 법, 규정 프레임워크에 대한 접근

European Commission 발표에 의하면, 유럽 집행위원회는 블록체인 기반의 응용프로그램과 관련된 분야에서 법적인 명확성과 또한 이를 위한 명확한 규제의 중요성을 인식하고 있다. EU는 법적 및 규제 분열을 피하기 위해 블록체인에 대한 EU 차원의 규정을 강력히 지지하고 있다. 이를 위해 위원회는 투자시장에서 소비자와 투자자를 보호하기 위해 암호화 자산 구제를 위한 포괄적인 입법 제안을 채택했다(ShapingEDF,

2015).

④ 디지털 유로(Digital Euro)

유럽중앙은행(ECB)과 유럽 집행위원회는 디지털 유로(Digital Euro)의 도입 가능성과 관련된 광범위한 정책, 법률 및 기술적 질문을 공동으로 검토하고 있다. 그들은 관련 조약에 규정된 각각의 명령과 독립성의 맥락에서 이것을 고려하고 있다.

⑤ 암호화 자산에 대한 새로운 EU 법률 제안 (EUC, 2020)

금융상품시장 지침에 따라 금융상품으로 인정되는 암호화 자산은 이전의 EU 증권시장 법률을 적용받고 있다.

그러나 이러한 규칙은 암호화 자산이 출현하기 이전의 것이었다. 이것은 기술의 혁신을 방해할 수 있다라는 입장을 취하고 있다. 그러므로 위원회는 암호화 자산 형태로 금융상품을 거래하고 해결하려는 시장 인프라에 대한 파일럿 체제를 제안했다(EUC 2020). 파일럿 체제는 기존의 규칙에 대한 면제를 허용하고 규제 기관과 회사가 블록체인을 활용하는 혁신적인 솔루션을 테스트하도록 하는 역할을 한다.

유틸리티 토큰 또는 지불 토큰과 같은 금융상품으로서 자격이 없는 암호화 자산의 경우, 위원회는 특정한 새로운 프레임워크를 제안했다. 이 프레임워크는 현재 암호화 자산의 발행, 거래 및 저장을 관리하는 모든 EU 및 국가 규칙을 대체하게 된다. 다시 말해 암호화 자산을 위한 별도의 새로운 규정을 만드는 것이다.

MiCA(Markets in Crypto-Assets Regulation)은 소비자 암호화 화폐 거래소의 무결성을 보호하면서 혁신을 지원(EUC, 2020)하도록 하고 있다. 여기에서 내부자 거래 혹은 프론트 러닝 금지와 같은 규칙이 포함된다. 제안된 규정은 암호화 자산을 발행하는 법인, 이러한 암호화 자산에 대한 서비스를 제공하는 회사, 디지털 지갑을 운영하는 회사들을 포함한다.

⑥ 범유럽 블록체인 규제 샌드박스

샌드박스는 규제 기관, 회사 및 기술전문가들이 모여 혁신적인 솔루션을 테스트하고 이를 배포할 때 발생하는 장애물들을 식별하는 것이다.

유럽 블록체인 파트너쉽(European Blockchain Partnership)은 데이터 이식성, 기업 간 데이터 공간, 스마트 계약 및 디지털 ID를 포함하여, 유럽 집행위원회와 협력하여 범유럽 규제 샌드박스를 계획하고 있다. 여기서는 헬스, 환경, 모바일리티, 에너지 등 다양한 분야에 대한 검토를 하고있다.

2) 미국의 블록체인 관련 법제도 현황

암호화 화폐와 기타 디지털 자산에 대한 소유와 판매를 위해서는 연방정부 법과 주 정부 법에서 복잡한 보완을 필요로 한다. 다음은 관련된 주요 법 영역이다.

① 세금 분야

세금 분야에서는 암호화 화폐 이용의 다음 케이스들이 주요 관심 영역이다.

- 암호화 화폐의 현금환전
- 암호화 화폐로 상품 혹은 서비스 구매
- 암호화 화폐를 다른 암호화 화폐로 교환

암호화 화폐를 채굴하는 자체가 income tax 대상이 된다. 만약 취미로 할 경우 기타 소득으로 처리하여야 한다. 이와 같이 미국은 납세 양식 1040의 관련 항목을 기재할 때 암호화 화폐에 대하여 어떻게 처리할지 구체적으로 가이드하고 있다. IRS Notice 2014-2021에서는 가상화폐를 자산으로 간주한다는 것을 명시하고 있다. 여기서 Notice는 법은 아니지만 법 집행에 관련된 공고라고 보면 합당하다. 이 Notice에서는 가상화폐 수입에 대한 납세뿐만 아니라 연말정산 시 가상화폐 손실로 인한 세금 환급 부분도 명시하고 있다. 이와 관련하여 납세자가 사용하여야 할 보고 양식들(IRS Form 1040, IRS Form 8949)도 세부적으로 제공하고 있다.

② 증권 분야

미국의 증권거래위원회(The Securities and Exchange Commission, SEC)는 투자 거래의 성격을 가진 모든 토큰이나 가상화폐에 대한 발행과 매매에 대한 규제 권한을 가지고 있다.

미국의 증권법(Securities Act) 2조 a항 1번과 증권거래법(Securities Exchange Act) 3조 a항 10번에 의하면, 증권은 투자 계약(Investment Contract)을 포함한다고 명시되어 있다. 여기서 투자 계약(Investment Contract)이란 US 대법원에서 정의하기를, 회사의 경영상 이익에 대한 상식적인 기대를 가지고 그 회사에 투자한 계약을 의미한다. 특정 토큰이 투자 계약인가 아닌가에 대한 판별을 위해서 SEC와 법원은 계약서의 양식보다 거래의 본질을 보는 것이다.

2019년 4월, SEC는 Framework for “Investment Contract” Analysis of Digital Assets 이란 웹사이트를 게시했다. 이 웹사이트에는 특정 디지털 자산이 투자 계약으로서 판매된 것인지 판단하기 위한 체크리스트가 제공되고 있다.

③ 상품 분야

2015년, 미국의 Commodities Futures Trading Commission(CFTC)은 가상화폐가 CFTC의 규제를 받는 상품(물자)이라고 규정했다. 이에 따라 The Commodities Exchange Act(CFA, 상품(물자)교환법)는 암호화 화폐 시장에서 사기와 조작에 대한 관할과 조사권을 CFTC에게 부여했다.

④ 자금 세탁

은행 보안법에 따라, 자금 서비스 사업(MSB, Money Service Business)은 연방정부의 자금세탁방지법에 의거한 규제 대상이 된다. 이에 따라 IRS(Internal Revenue Service)는 자금 서비스사업에 대해 감독권을 가지게 되었다. 흔히 이야기하는 자금 전송자(Money Transmitter)가 MSB의 한 예이다. 2013년 Financial Crime Enforcement Network은 이러한 자금 전송자에 가상화폐 교환과 가상화폐의 중앙 저장소 관리자를 포함 시켰다. 가상화폐를 이용한 자금 세탁을 제재한다는 것이다.

⑤ 기타

이외에도 각 주 정부는 가상화폐와 관련된 신탁, 자산, 공증 분야에서 각각의 법령들을 각자 마련하고 있다. 이외에도 많은 분야에서 가상화폐를 기존 화폐와 같이 투명하고 안전하게 관리하고자 하는 노력이 일찍부터 진행되고 있다.

제5장 결 론

오늘날 우리 사회에서의 SW 비중은 나날이 팽창하고 있다. 그와 더불어 SW 안전에 관심도 나날이 증가하고 있다. SW 비중이 커지는 요인 중 하나는 바로 SW신기술이 등장하면서 다양한 분야에서 혁신적인 활용이 이루어지고 있기 때문이다. AI, 메타버스, 블록체인이 그 대표적인 예이다.

SW 안전은 전통적으로 SW 공학 분야에서 연구되어왔으며, SW 안전 분석에 관련된 기술들이 주로 연구되어왔다. SW 공학 분야에서 연구되어 온 안전 기술은 주로 SW에 특화된 결함 트리 분석과 같은 준형식적 기법(Semi-Formal Method)과 시계열적인 논리체계와 상태차트 등을 활용하는 형식적 기법(Formal Method)이 주를 이루어 연구되어왔다. 실용적 관점에서 바라보면 이러한 준형식적 기법은 이미 알려진 위험원이 있다고 가정하고 이러한 위험원이 발생할 조건들을 분석하는 접근이다. 반면 형식적 기법는 철저한 분석 관점에서 접근하는 방법이다.

철저한 분석을 하는 형식적 기법은 분석 결과의 이해도와 설득력 측면에서 이론적으로 강점이 있으나, 상태의 발산적 다양성과 같은 현실적으로 거의 무한한 문제 공간을 가진다는 어려움이 있다. 따라서 형식적 기법은 SW 시스템에서 매우 국한된 모듈에 대하여 적용하는 정도가 현실적인 활용이라고 볼 수 있다. 특히 소스 코드가 억단위 라인수로 늘어나고 있는 융합 시스템에서는 더욱 그러하다.

현실적으로 산업 도메인에서는 철저한한 접근보다 알려진 위험원에 대한 예방적인 차원에서 안전 메커니즘을 구현하는 접근이 선호되고 있다고 볼 수 있다. 이미 많이 알려진 기능 안전(Functional Safety)표준들이 그렇다. 자동차, 철도, 항공 등에서 이미 잘 정책화되어 사용되고 있는 기능 안전은 도메인 별로 특화된 표준들(예, 자동차 전장 시스템 분야의 ISO 26262)이 이에 속한다. 이러한 기능 안전 표준에서는 시스템 개발의 초기 단계에서 잠재 위해들을 식별하고 이를 위한 안전 요건들을 시스템, HW, 그리고 SW상에서 구현하는 방식을 취하고 있다. 기능 안전에서의 핵심 성공 요인은 바로 잠재적인 위험원(Potential Hazard)을 얼마나 빠짐없이 식별하는가에 달려 있다. 산업 도메인 별로 잠재적인 위험원들을 식별하는 것은 오랜 시간 동안 축적된 노하우들에 달려 있다고 볼 수 있다. 특정 시스템의 잠재적인 위험원을 식별하기 위해서는 해당 시스템을 오랜 기간 개발해온 도메인 전문가가 필요하다. 이러한 도메인 전문가들은 해당 시스템 구현에 사용되는 기술들과 사용 환경상에서의 위험 요건들에 대한

축적된 전문지식을 가지고 있다고 볼 수 있다.

SW 신기술의 경우에도 이론적으로 동일한 접근이 필요하다. 신기술 자체에 대한 잠재 위험들과 신기술이 사용되는 도메인 특화 잠재적 위험원들이 식별되고 축적되어 새로운 시스템을 개발할 때 개발 초기 단계에 잠재 위험들에 대한 안전 요건들이 도출되고 안전 메커니즘이 구현되어야 한다. 그러나 SW 신기술의 경우 이러한 잠재 위험들에 대한 축적된 데이터 자체가 부재하다고 볼 수 있으며, 동시에 기능 안전 분야와 같이 안전에 대한 제도화 자체도 부재한 상황이다. 그럼에도 불구하고 SW신기술은 진흥의 명목으로 지속적으로 확대 사용되고 있다.

본 연구는 SW신기술 분야 안전에 대한 기초 연구라고 볼 수 있다. 본 연구를 시작으로 SW신기술 분야의 안전기술에 대한 지속적인 연구와 잠재 위험에 대한 지속적인 축적에 체계적으로 이루어져야 할 필요가 있다.

본 연구는 SW신기술에 대한 사고조사와 이러한 사고 들의 잠재 원인을 추정하고 이를 기반으로 위험 유형을 가급적 유형화 하는 것에 집중했다. SW 신기술에 대한 조사는 SW 안전 전문가의 정성 연구와 언론기사 빅데이터 분석을 통해 총 37건의 사고를 선별했다. 원론적으로 접근한다면, 샘플링된 각 사고들에 대한 사고원인 분석을 진행하고 이들을 유형화하여야 하는데, 이를 위해서는 각 사고 시스템에 대한 전문적인 기술자료(사양서, 설계서, 소스 코드, 검증 데이터 등)들이 필요하다. 현실적으로 이러한 데이터를 확보하는 것은 불가능하고, 또한 확보한다고 하여도 통상적으로 단일 시스템에 대한 원인 분석 작업이 1년에서 수년이 걸리므로, 현실적으로 불가능한 접근이라고 볼 수 있다. 유사한 문제가 기존의 기능 안전 분야에서도 수십년전 초기 단계에도 있었으리라 추측된다.

이러한 제약 조건상에서 현실적으로 필요한 접근은 잠재 사고원인에 대한 이론적인 위험원 유형을 먼저 만들어 보고 이를 기반으로 각 사고들에 대하여 전문가적 견해로 잠재 원인을 추측하여 보는 것이다. 이와 같은 접근을 기초로 하여 향후 연구들이 보다 체계적으로 진행될 수 있을 것이라 판단된다.

본 연구에서는 먼저 이론적인 잠재 위험원 유형 모델을 먼저 개발했다. 이론적인 초석을 확보하기 위하여 잠재 위험원 유형 모델은 기존에 SW 공학 분야에서 결함 및 실패 모델들을 조사하여 그중 가장 적합한 모델을 선택하고 1차적인 모델을 개발하였다. 1차 모델을 기반으로 SW신기술과 안전 공학에서 사용하는 안전 메커니즘의 컨셉을 반영하여 전문가들이 모여 2차 모델로 확장하였다. 2차 모델까지가 시스템 내부와 사용자 특성의 일부분을 반영한 모델이다.

2차 모델을 기반으로 사고 데이터에 대한 사고원인 추정을 시도했는데, 언론에 노출된 정보를 기반으로 사고원인 자체를 기술적으로 추정하기에는 어려움이 많았으며, 또한 언론에서 언급한 사고원인들의 상당수가 시스템 경계내의 요소 외에도 환경적인 요소들을 많이 다루고 있었다. 이 부분에서 사고원인에 대한 추정은 보다 포괄적이고 거시적으로 이루어져야 할 필요성을 확보하게 되었다.

사고원인을 포괄적으로 유형화하기 위해서는 사고원인이 존재하는 영역에 대한 구분이 먼저 이루어져야 한다. 시스템 공학적으로 이 영역은 시스템, 사용자, 그리고 환경 3가지의 영역으로 구분될 수 있다. 실제 안전 공학 전문가인 Nancy Levison 교수의 STPA(Systems-Theoretic Process Analysis) 방법론에서도 유사한 영역으로 잠재 위험군을 바라보고 있다. 그러므로 이러한 분류는 매우 합당하다고 볼 수 있다. 다만 STPA도 잠재 사고원인에 대한 유형을 분류하는 것에는 세부적인 가이드라인을 제공하고 있지는 않고 위험 분석 기술에 초점을 맞추고 있다.

본 연구에서는 잠재 사고원인에 대한 유형을 개발하는 것에 초점을 맞추고자 보다 거시적으로 안전사고 원인 조사 방법론을 조사하여 보았는데, 그 중 본 연구의 의도와 가장 부합하는 방법론으로 영국의 ICAM(Incident Cause Analysis Method)을 활용하여 잠재 사고원인에 대한 유형을 포괄적으로 확장하도록 하였다. ICAM은 구조화된 프레임워크를 기반으로 사고원인 조사를 실시하는 체계적인 방법론으로서 영국과 호주 등, 다양한 국가의 다양한 산업 도메인에서 사용되고 있다.

이러한 과정을 거쳐 사고원인 유형을 분류하는 SW 사고 유발 위험원 프레임워크를 개발하였다. SW 사고 유발 위험원 프레임워크는 상위 7개의 위험원 분류로 시작된다. 1. Safety Mechanism, 2. Actor/Human, 3. Environment, 4. Organization /Process/Policy, 5. SW system, 6. Data, 그리고 7. HW(센서 및 액츄에이터) 가 그 상위 분류 유형이다. 각 상위 분류는 하위 분류를 갖게 된다. 여기서의 7개 분류를 System Theory의 세 개 영역으로 그룹핑 하면 다음과 같다. 첫 번째 환경 분야에는 Environment와 Organization /Process/Policy가 속한다. 두 번째 사용자 분야에는 Actor/Human이 속한다. 마지막으로 세 번째 시스템 분야에는 나머지 Safety Mechanism, SW system, Data, 그리고 HW가 속한다.

유형 분류의 편의를 위하여 각각의 세부 유형들에 대한 설명을 제공하였다. 이러한 유형 프레임워크를 기반으로 37개 정리된 사고 데이터에 대하여 전문가들의 원인 추정 의견서를 작성하도록 하였고, 전문가들은 추정 가능한 주요 원인들에 대하여 각 사고 별로 의견을 제시하였다. 추정된 의견을 기반으로 주요 잠재 원인을 정리하면 다음

과 같다.

가장 많은 원인으로 거론된 유형이 안전장치 부재(Safety Mechanism 부재)이다. 이는 곧 잠재 위험원에 대한 분석이 이루어지지 않았다고 추론할 수 있고(위험 관리 미흡), 많은 경우 그 이유는 안전 제도의 미흡이 주요 원인으로 추정되었다(안전 제도 미흡). 그 외 특이한 사항으로는 AI, 메타버스, 블록체인을 불문하고 의도적으로 사고를 내는 경우(악의적 사용)가 많았다. 블록체인의 경우 예상한 바와 같이 보안 이슈(Security 오류)가 많았다. AI의 경우 두드러지는 잠재 원인으로서 데이터에 대한 불충분과 부적합 이슈가 많았다. 또한 머신러닝의 경우 학습데이터의 성격에 따라 편향되는 이슈(데이터 편향)도 두드러졌다. AI의 경우 사용자 기능 오류가 높은 이유는 학습 자체가 올바르게 안되다 보니, 당연히 기능에 오류가 생기는 것을 의미한다. 그 외 사용과 운영에 있어서 잠재 위험을 관리하는 절차(사용/운영 절차 미흡)에 대한 부족도 두드러졌다.

마지막으로 SW 신기술 안전 확보를 위한 글로벌 표준, 정책, 법/제도의 현황 조사를 진행하였는데, 아직까지는 구체적인 정책 및 법/제도 존재하지는 않았다. 다만 AI과 같은 SW에 대하여 제조물 책임법과 같은 안전에 대한 책임부가 관련 법 제도에 대한 상정들이 이루어지고 있는 상태다. 또한 처벌에 관련된 형사법의 경우, 각 사회 분야의 현행법을 기준으로 그대로 적용하는 접근을 취하고 있다. 안전 기술에 관하여서는 SW 신기술 분야에 특별히 따로 개발하여 사용하는 것은 아직까지는 없는 것으로 보인다. 이러한 현상은 선진국의 일반적인 SW에 대한 경우 현행법/제도, 표준들이 상대적으로 잘 정책화되어 있기에 서두르지 않는 것으로 판단된다. 반면에 우리나라의 경우 이러한 SW 분야의 법제도 및 표준이 아직 실무에 정책화되어 있다고 볼 수는 없다. 하나의 대표적인 예로서, 비기능 요구사항, 특히 안전에 관련된 요구사항에 대한 식별과 반영 등이 거의 부재하다고 볼 수 있다.

본 연구에서 제시된 잠재 위험원 유형은 안전사고, 사회 사고의 원인을 조사하는 기본적인 방법론의 시작점으로 활용될 수 있을 것으로 보인다. 그동안의 국내 많은 사고들을 볼 때, 대부분 그 원인이 인재, 혹은 매뉴얼 부족, 장비 오류 정도의 모호한 유형으로 결론 나는 경우가 많았다고 볼 수 있다. 이와 반대로 선진국의 경우 세부적인 원인 분석이 일어나고 이에 대한 공유가 되고, 재발 방지를 위한 세부적인 조치가 진행된다.

본 연구에서 제시된 위험원 유형은 아직 초안 단계로서 보다 고도화 될 필요가 있다고 판단된다. 그러나 안전사고 원인 분석의 초석으로서 그 가치가 충분하다고 사료 된다. 추가적으로 필요한 것은 ICAM과 같은 안전사고 원인 분석 방법론이다. 우리나라

에서 어떤 안전사고 원인 분석 방법론을 사용하는지는 파악하기가 매우 어렵고, 실제로 존재하는지도 불분명하다. 많은 경우 전통적인 재난관리에서 사용하는 접근을 사용하는 것이 아닌가 추정된다. 이러한 전통적인 재난관리 방법론은 전통적 사회에서의 자연재해에는 도움이 될 수 있으나, 첨단 사회의 사회재난에는 한계가 있으리라 판단된다.

이를 위하여 본격적으로 안전사고 원인 분석 방법론을 개발하고 각 부처에 보급하며, 실제 SW 안전사고 조사위원회를 구성하고 운영, 지원하는 실행 계획이 시급하다고 판단된다.

[별첨] SW 신기술 사고 조사 보고서 목록

SW신기술 안전사고 보고서

No. AI-01

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">신약 개발 AI의 악용 가능성 제기</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hani.co.kr/arti/science/science_general/1037549.html</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 4월 5일</td></tr>
<tr><td>작성자</td><td colspan="3">곽노필</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>AI의 악용</td></tr>
<tr><td>사고 일시</td><td></td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 AI 기술
인류가 인공지능에 거는 기대 가운데 큰 몫을 차지하는 것이 신약 개발 분야로, AI는 제약 산업에서도 다방면으로 활용하고 있다.
신약을 개발하려면 수천, 수만 가지의 화학물질을 탐색하고 조합해 후보 물질을 찾아내는 단계를 거쳐야 하는데 인공지능의 빠르고 정확한 데이터 조합 및 분석 능력을 활용하면 이 시행착오 기간을 크게 단축할 수 있다. 최근 AI 신약개발지원센터장에 취임한 김우연 카이스트 교수는 기자간담회에서, 인공지능을 활용하면 신약 개발기간을 15년에서 7년으로 줄일 수 있을 것이라는 기대를 표명한 바 있다.

2. 실험 내용
1) 실험 목적
AI 알고리즘이 긍정적인 작업이 아닌 부정적인 작업으로 설정되어 있다면 얼마나 빠르고 쉽게 AI 알고리즘을 남용할 수 있는지 확인하기 위해 이번 실험이 진행되었다.
컬래보레이션스제약(Collaborations Pharmaceuticals)이라는 미국의 신약 개발업체 연구진에 의해 진행된 이 연구는 스위스의 슈피츠연구소(핵과 생화학무기방어를 위한 스위스연방연구소)가 2년마다 여는 생화학무기회의에 신약 개발 인공지능의 오용 가능성을 보여달라는 요청에 따라 이뤄졌다.
이 회사 인공지능 연구진의 본래 업무는 신약 물질 분자 설계에 유용하도록 치료 및 독성을 예측할 수 있는 기계학습 모델을 구축하는 것으로 치료 물</td></tr>
</table>

질이 생명에 필수적인 다양한 종류의 단백질에 개입할 수 있는 위험, 즉 독성을 피할 수 있도록 하는 것이 이들의 목표다.
신약 물질을 물리적으로 실험하기에 앞서, 분자 설계를 주로 하는 이 연구진은 “수십 년 동안 우리는 인간의 건강 악화가 아닌 개선을 위해 컴퓨터와 AI를 사용해 왔다”면서 이 기술이 오용될 가능성에 대해 순진하고 안일했음을 반성했다.

2) 실험 방법
연구진은 적절한 약물 후보를 탐색하는 데 사용하던 ‘메가신(MegaSyn)’이라는 머신러닝 알고리즘을 조금 변형하여 써 보기로 했다. 원래 메가신은 일반적으로 물질의 독성이 예측될 경우 감점하고, 반대로 생체적합성이 예측되면 점수를 더하는 식으로 학습하는 AI였다. 하지만 이번 연구에서는 독성과 생체적합성 모두에 가점하는 방식으로 알고리즘을 바꾸었다. 즉, 의도적으로 인체에 치명적인 작용할 독성 물질을 찾고자 한 것이다. 연구진은 누구나 접근할 수 있는 공개 데이터베이스에 있는 정보로 AI를 훈련 시켰으며, 살충제, 환경 독소 및 기타 약물로 구성된 반수치사량(LD50) 모델을 기준 삼아 새로운 가상 분자에 점수를 매겼다.

3) 실험 결과
실험 결과 AI는 6시간 만에 화학 무기로 사용될 수 있는 4만 종의 독성 물질을 찾아냈고 그 중에는 신경가스 'VX'를 능가하는 독성도 포함된 것으로 알려졌다. VX는 화학무기금지협약 상 화학 무기로 분류되는 물질로, 김정은의 형 김정남 암살에 사용된 가장 강력한 신경 작용제이기도 하다. 뿐만 아니라 기존 화학 무기와 유사하지 않지만 VX 보다 독성이 강한 물질도 확인됐다.
살충제, 환경 독소, 약물 데이터로부터 이처럼 강력한 신종 독성 물질이 고안된 것은 실험진도 예상하지 못한 부분이었다.

3. 시사점
이번 실험은 누군가가 실행에 옮길 경우 완전히 새로운 종류의 화학 무기가 개발될 수 있음을 시사한다.
단지 기계학습 모델의 훈련 방법을 살짝 바꿨을 뿐인데 인공지능은 약물을 개발하는 이로운 도구에서 치명적 분자 생성기로 변신했다. 연구진은 “우리가 개발한 독성 모델은 원래 독성을 피하는 데 사용하기 위해 만들어진 것인데, 실험 결과 독성 예측력이 좋을수록 치명적인 분자 설계 능력도 좋아졌다”고 밝혔다.
연구진은 특히 이런 식의 인공지능 오용에 대한 진입 장벽이 너무 낮은 현실에 우려를 표명했다. 이번 연구를 이끈 컬래보레이션스제약의 수석과학자 파비오 어비나(Fabio Urbina)는 미국의 온라인매체 ‘더 버지’ 인터뷰에서 “온라인에는 무료로 얻을 수 있는 생성모델과 독성 데이터 세트가 많다”며

	“파이썬 코딩법을 알고 기계학습 모델을 만들 수 있는 능력만 있다면 주말을 이용해 독성 데이터 세트로 구동되는 이런 생성 모델을 구축할 수 있을 것”이라고 말했다. 그는 이어 “우리는 나쁜 배우들에게 아이디어를 주고 싶지 않았기 때문에 이 논문을 발표할지 망설였다”며 “그러나 우리가 할 수 있다면 어딘가에 있는 누군가도 이미 그것을 생각하고 있거나 장차 생각할 가능성이 높다는 판단 때문에 경종을 울리기 위해 발표하기로 했다”고 덧붙였다.
사고 피해 내역(예상)	1. **피해 대상**: 전 인류가 될 가능성 2. **피해 내역**: 신약 개발 AI가 악용될 경우 사람들을 위험에 빠뜨릴 가능성이 있음

SW신기술 안전사고 보고서

No. AI-02

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">AI 딥페이크 기술을 이용한 여성 나체사진 유포</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">http://www.aitimes.com/news/articleView.html?idxno=132983</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2020년 10월 21일</td></tr>
<tr><td>작성자</td><td colspan="3">윤영주</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>텔레그램</td><td>사고 유형</td><td>딥페이크 범죄</td></tr>
<tr><td>사고 일시</td><td>2019년 7월 ~</td><td>사고 지속시간</td><td>1년</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 AI 기술
‘딥페이크’란 인간처럼 판단할 수 있도록 컴퓨터를 교육시키는 ‘딥러닝(deep learning)’과 가짜를 뜻하는 ‘페이크(fake)’의 합성어로 간단하게 말하면 AI를 이용해서 사람의 이미지를 합성하는 기술이다. 일반적인 AI 기술은 새로운 무엇을 만들어 내는 것이 아니라 다량의 데이터에 흐르는 공통된 특징을 찾아낸 뒤 그 결과에 따라 어떤 작업을 하게 되는데, 딥페이크에 사용되는 AI 기술은 한 걸음 더 나아가 스스로 무언가를 새로 만들어 낸다는 점에서 차이가 있다. 딥페이크 기술을 엔터테인먼트 분야에 활용 시 제작비 절감으로 이어져 해당 산업의 활성화에 기여하며, 글로벌 홍보를 위해 사용 시 지역과 장벽을 허물 수 있는 장점이 있다. 또 교육 분야에 활용하면 학습에 생동감을 더하게 된다.

2. 사고 경위
1) 보도 내용
영국 공영방송 BBC를 통해 공개된 사이버 보안업체 센시티(Sensity)의 보고서를 통해 인공지능(AI) 기술을 악용해 사진 속 여성들의 옷을 벗겨 누드로 바꿔주는 텔레그램 채널이 운영되고 있는 것으로 드러났다.
보고서에 따르면 2019년 7월부터 1년간 가짜 나체사진 유포로 피해를 입은 여성의 수는 약 10만 4852명으로 파악됐다. 조사 결과 미성년자 사진들도 발견돼, 일부 사용자들은 주로 소아성애 콘텐츠를 만들어 공유하는 데 이 서비스를 이용하고 있는 것으로 나타났다.

2) 범행 내용
문제의 서비스를 운영하는 텔레그램 채널은 AI 기반 ‘딥페이크 봇(deepfake bot)’을 이용한 것으로 알려졌다. AI 기술로 여성 이미지에서 옷을 지우고 나체로 합성하는 방식이다. 이 AI 봇은 텔레그램 사용자가 보낸 사진을 무료로</td></tr>
</table>

	몇 분 안에 뚝딱 편집해준다. 사용자가 자동화된 서비스를 통해 옷을 입은 여성의 사진을 익명으로 보내면 옷을 벗은 누드 버전의 사진을 받아볼 수 있다. 센시티는 해당 서비스 봇이 특히 러시아 소셜미디어 사이트인 'VK'를 통해 상당한 광고를 받았으며 사용자의 약 70%가 러시아와 구소련 국가 출신이었다고 말했다. 그러나 VK 측은 "플랫폼에서 그와 같은 콘텐츠나 링크는 허용되지 않으며 이를 배포하는 커뮤니티를 차단하고 있다"고 선을 그었다. 센시티 측은 이번 보고서 내용을 텔레그램을 비롯해 관련 법 집행기관 등에 알렸으나 아무런 답변도 받지 못했다고 설명했다. BBC도 텔레그램 측에 의견을 요청했으나 이번 논란과 관련해 어떠한 언급도 하지 않았다고 전했다. 3) 채널 운영자 입장 이 텔레그램 채널의 운영자는 이 같은 서비스를 단순한 장난 혹은 오락거리로 치부해 공분을 샀다. 'P'로 알려진 해당 서비스 운영자는 "폭력성 없는 오락거리일 뿐"이라며 "사진도 사실적이지 않기 때문에 아무도 이 사진을 이용해 협박하려고 하지 않을 것"이라고 주장하고 이어 "운영팀이 어떤 사진이 공유되고 있는지 모니터링하고 있다"며 부적절한 사용자들은 차단시키고 있다고 설명했다. **3. 문제 제기** 지금까지 딥페이크를 악용한 피해 사례들을 살펴보면 유명인사의 가짜 포르노 영상 등이 주로 거론되었지만 이제 딥페이크 범죄 대상이 일반인들로까지 확대되고 있는 양상이다. 실제 이번 피해자의 사진 가운데 약 70%가 소셜미디어 계정이나 개인 자료에서 나온 사진이었다. 딥페이크 나체사진 공유 채널 중 한 곳에서 실시한 설문조사에 따르면 사용자 대부분은 그들이 실생활에서 알고 있는 여성의 사진을 나체 이미지로 만들고 싶다고 답했다. 조르지오 파트리니 센시티 CEO는 민간인의 개인 사진이 도용됐다는 점을 들어 "사진이 공개된 소셜미디어 SNS 계정만 있다면 누구나 표적이 될 수 있다"고 경고했다. 또 "이 같은 웹사이트나 앱 상당수는 법적으로 엄격히 규제되지 않고 있다"며 문제의 서비스가 공공연하게 운영되고 있는 현실을 꼬집었다. 특히 딥페이크는 악용되는 목적이나 용도가 다양하기 때문에 적발·예방·근절이 어렵다. 전문가들은 장기적으로 딥페이크 오디오·비디오 등의 가짜 콘텐츠 양산이 오디오·시각 자료에 대한 광범위한 불신으로 이어질 수 있다고 경고했다.
사고 피해 내역	**1. 피해 대상:** 약 10만 4852명의 여성, 일부는 미성년자로 확인됨 **2. 피해 내역:** 텔레그램 채널을 통해 가짜 나체사진이 유포됨

SW신기술 안전사고 보고서

No. AI-03

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">자율주행차량에 의한 첫 보행자 사망사고</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hankyung.com/it/article/2018032081821</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2018년 3월 20일</td></tr>
<tr><td>작성자</td><td colspan="3">송형석</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국(애리조나)</td><td>사고 유형</td><td>사망사고</td></tr>
<tr><td>사고 일시</td><td>2018년 3월 18일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 자율주행차와 AI
자율주행자동차란 자율주행을 위해 자동차에 IT·센서 등 첨단 기술을 융합하여 스스로 주변 환경을 인식, 위험을 판단하고 주행 경로를 계획하여 운전자 또는 승객의 조작 없이 안전한 운행이 가능하도록 한 자동차를 말한다.
자율주행차는 기본적으로 도로 위의 사물의 유무를 판단하고 그것을 넘어 단순 장애물인지 사람인지까지 판단할 수 있어야 한다. 자율주행차가 주변 상황을 판단하기 위해서는 다양한 기술이 필요한데 여기서 필요한 기술이 인지(주변 상황 감지), 판단(자동차가 어떤 동작을 취해야 할지 결정), 제어(동작 결정 후 최종적으로 차를 움직이게 조작), 측위(차가 도로의 어디에 있는지를 알아내는 기술) 등이 있다.
자율주행차에서 AI는 판단 기술에 적용되는데 인지 기술로 얻어낸 정보를 바탕으로 분석해, 자동차가 어떤 동작을 취해야 하는지 결정을 하게 만든다.

2. 사고 경위
1) 사고 발생
2018년 3월 18일 밤 미국 애리조나 템피에서 자율주행 시험 중이던 우버 차량이 횡단보도가 아닌 곳에서 자전거를 끌고 도로를 건너던 여성을 치여 숨지게 한 사고가 발생했다. 이는 자율주행차 최초의 보행자 사망사고로 기록됐다.
당시 우버는 스웨덴 볼보자동차가 생산한 SUV 'XC90'로 시험주행을 하였으며, 운전석에 운전자가 앉아 있었지만, 시속 64km의 3단계 자율주행 모드로 운행 중이었다. SAE에 따르면 3단계는 ‘조건부 자율주행’으로 도로나 날씨 조건 등 자율주행이 가능한 환경에선 운전자가 ‘전방 주시 의무’ 없이 책을 보거나 옆 사람과 대화를 해도 알아서 갈 수 있는 수준이다. 다만 운전자 개입이 필요하다고 판단될 때는 경고와 함께 시스템 제어권이 넘어오기 때문에 운전자는 언제든 신속하게 대응할 수 있게끔 주의 의무를 기울여야 한다.</td></tr>
</table>

	2) 사고 원인 조사 사고 직후 이뤄진 미 연방교통안전위원회(NTSB) 조사에 따르면, 비상브레이크 시스템이 제대로 작동하지 않았기 때문이라고 사고 원인을 밝혔다. 2018년 5월에 나온 NTSB 예비조사 보고서에는 사고 차량에 탑재된 자율주행 센서가 충돌 6초 전쯤 보행자를 감지했지만, 소프트웨어가 이를 차량이나 자전거로 잘못 식별했고 운전자를 향한 경고도 이뤄지지 않았다는 내용이 담겼다. 게다가 우버 엔지니어들이 시스템 오작동을 줄이기 위해 자율주행 시스템과 연동되지 않은 볼보 자동차의 자체 자동 비상브레이크 시스템을 비활성 시켜 놓은 점 역시 사고 예방을 막았다. NTSB는 “비상브레이크 시스템이 제대로 작동했을 경우 약 1.3초 전 충돌을 피할 수 있었을 것”이라고 했다. 3) 사고에 대한 우버 측 입장 우버는 해당 사고 발생 직후, 소셜미디어를 통해 숨진 보행자에 대한 애도를 표하고 경찰 수사에 최대한 협조하고 있다고 밝혔다. 또한 우버는 사고가 발생하자 애리조나주 피닉스·템페와 피츠버그, 샌프란시스코, 토론토 등지에서 진행하던 자율주행차 시험 운행을 전면 중단했으며, 약 9개월만인 2018년 12월에 피츠버그에서 자율주행차 시험주행을 재개했다. 시험주행을 재개하면서 에릭 메이호퍼 우버 ATG 대표는 성명을 통해 “지난 9개월간 안전의 핵심을 강화하고자 온갖 노력을 기울이며 수많은 검사에 많은 시간을 들였다”고 말했다. 3. 수사 결과 미국 검찰과 경찰은 자율주행차에 의한 세계 첫 보행자 사망사고에 대해 자율주행 시스템이 아닌 운전자가 사고 책임을 져야 한다는 판단을 내놨다. 앨리스터 아델 담당 지방 검사는 기소 이유에 대해 “운전자가 운전석에 앉았다면 그 차량을 안전하게 통제하고 운전할 책임이 있다”고 말했다. 자율주행차라도 운전석에 앉았다면 사고 책임을 피할 수 없다는 것이다. 실제 템피 경찰 조사에서 테스트 차량 운전자는 충돌 당시 스마트폰으로 스트리밍 방송을 보고 있던 것으로 나타났다. 차량 데이터를 조사한 결과 이 운전자는 충돌 1초 전에야 운전대를 잡았고, 충돌 직후 브레이크를 밟았다. 경찰은 “(운전자가 주의했다면) 이번 충돌은 전적으로 피할 수 있었다”고 했다. NTSB 역시 2019년 11월 “운전자가 주행 중 개인 휴대전화를 보느라고 주의가 산만해졌다”며 운전자의 부주의를 사고의 직접적인 원인으로 봤다. 미국자동차공학회(SAE) 기준상 주행 보조로 분류되는 1~2단계와 달리 3단계 이상 자율주행은 자동차가 주도적으로 운전을 하는 만큼 사고 책임 역시 시스템 또는 자동차 제조사에 물어야 하는 것 아니냐는 세간의 인식과 다른 판단을 한 것이다.
사고 피해 내역	**1. 피해 대상:** 2018년 3월 18일 밤 미국 애리조나에서 횡단보도가 아닌 곳에서 자전거를 끌고 도로를 건너던 여성 **2. 피해 내역:** 자율주행차량의 첫 보행자 사망사고

SW신기술 안전사고 보고서

No. AI-04

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">숨진 가족 목소리 되살리는 아마존 AI '알렉사'의 신기능 논란</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.yna.co.kr/view/AKR20220624003600075</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 6월 24일</td></tr>
<tr><td>작성자</td><td colspan="3">정윤섭</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>AI 음성 복제 기능 악용 가능성</td></tr>
<tr><td>사고 일시</td><td>2022년 6월 22일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 플랫폼
1) AI 비서
AI 비서는 사용자의 음성을 인식해 사용자의 명령에 따라 음악을 재생하거나 스마트폰, 가전제품이나 자동차 등 IoT로 제어할 수 있는 기기 등을 작동하는데 사용된다.
원조는 애플사의 시리(Siri)인데 초기에는 인식률이 낮아 활용도가 높지 않았지만, 업그레이드를 거듭하면서 기능이 한층 강화됐다. 2014년에 아마존에서 알렉사가 출시되면서 AI 비서 시장의 경쟁이 본격화되기 시작했다

2) 알렉사(Alexa)
알렉사는 미국 최대 전자상거래업체인 아마존에서 2014년에 11월에 출시한 음성인식 인공지능(AI) 비서이다. 아마존에서 원통형 스피커 ‘에코’를 사서 설치하면 목소리로 각종 가전기기나 난방, 조명 등을 작동할 수 있고, 자동차와 연계해 추운 날 집안에서 시동을 걸고 히터를 미리 켜둘 수도 있으며 궁금한 뉴스를 알려주거나 일상적인 팁을 제공하기도 한다.
원래 인터넷에 익숙하지 않은 사람과 대화하며 필요한 물건을 살 수 있도록 전자상거래를 돕는 AI 비서로 개발되었으나, 그 쓰임새가 점점 확대되고 있다, 알렉사는 클라우드 기반으로 작동되기 때문에, 알렉사를 자주 사용할수록 사용자가 말하는 패턴, 단어, 개인적인 기호 등을 학습해서 더 잘 받아들인다.

2. 알렉사의 신기능 소개
2022년 6월 22일 미국 네바다주 라스베이거스에 열린 아마존 리마스(Re:Mars) 컨퍼런스에서 AI 음성 비서 ‘알렉사’의 새로운 음성 복제 기능이 공개되었다.
이날 행사에서 시연된 기술은 한 어린이가 알렉사를 향해 ‘할머니 목소리로</td></tr>
</table>

	오즈의 마법사 책을 읽어달라'고 요청하자 알렉사는 할머니 음성으로 전환해 책을 낭독했는데, 이때 알렉사가 모방한 할머니 목소리는 세상을 떠난 고인의 목소리였다. 알렉사의 이 신기능은 1분 미만의 음성 샘플만 있으면 특정인 목소리를 그대로 복제해 흉내 낼 수 있고, 살아있는 사람뿐만 아니라 숨진 가족의 목소리까지도 재현해낼 수 있다. 아마존 측은 알렉사의 음성 모방 기능이 언제 출시될 것인지에 대해서는 언급하지 않았다. **3. 아마존 측 입장** 기술 시연을 진행한 아마존의 로히트 프라사드 수석부사장은 알렉사에 공감이라는 인간적 속성을 더 부여했다고 설명하면서"AI가 코로나19로 사랑하는 가족을 잃은 고통을 없애주지는 못하지만, 가족과의 추억을 오래 간직할 수 있다"말했다. 또한, "앞으로 더 짧은 분량의 음성 데이터를 바탕으로 고품질의 목소리를 재현해내는 방안을 개발할 것"이라고 설명했다. **4. 기술에 대한 반응** 알렉사의 신기능에 대해 미국의 워싱턴포스트(WP)는 알렉사의 신기능이 윤리적인 문제와 더불어 사망자의 개인정보 권한을 둘러싼 논란을 초래한다고 논평하며 전문가 의견을 첨부했다. 온라인 보안업체 소셜프루프시큐리티의 레이철 토백 최고경영자(CEO)는 "세상은 사용자 친화적인 음성 복제 기술을 받아들일 준비가 되지 않았다"며 알렉사의 신기능이 사이버 범죄에 악용될 수 있다고 지적했다. 또한 "사이버 범죄자가 음성 샘플을 사용해 다른 사람 목소리를 복제해낸다면 이는 사기와 데이터 탈취, 계정 도용 등으로 이어질 수 있다"고 경고했다. 호주 커틴대학의 타마 리버 인터넷학 교수는 숨진 사람의 목소리를 흉내 내는 기술이 사람과 기계 사이의 경계를 모호하게 만드는 위험이 있다며 "죽은 사람의 데이터를 사용하는 것은 섬뜩하고 비윤리적"이라고 비판했다. **5. 시사점** AI의 목소리 모방 기술은 미디어·콘텐츠 분야에서도 활용도가 매우 높을 것으로 기대받고 있으며, 실제로 고인을 소재로 만드는 AI 콘텐츠가 더욱 늘어나고 있다. 하지만 AI 분야 전문가들은 AI의 음성 복제 기술의 시장 가능성에 대해 인정하면서도 동시에 부작용 역시 클 것으로 예상하고 있다. 저작권, 인권 침해, 범죄 악용 등의 악용 사례가 증가한다면 진보된 기술은 오히려 우리에게 큰 혼란을 일으키는 양날의 검이 될 것 이라고 예상하는 것이다.
사고 피해 내역(예상)	**1. 피해 대상**: 불특정 다수가 될 가능성이 있음 **2. 피해 내역**

	(1) 범죄자가 복제된 음성 샘플로 사기와 데이터 탈취, 계정 도용 등 사이버 범죄에 악용할 가능성 (2) 고인의 음성 데이터에 대한 소유권 분쟁 초래 가능성

SW신기술 안전사고 보고서

No. AI-05

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">AI 챗봇 '이루다' 의 혐오 발언 및 개인정보 유출로 인한 서비스 중단</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">http://www.aitimes.com/news/articleView.html?idxno=135579</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 1월 12일</td></tr>
<tr><td>작성자</td><td colspan="3">박성은</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>한국(이루다)</td><td>사고 유형</td><td>개인정보 유출 외</td></tr>
<tr><td>사고 일시</td><td>2020.12.~ 2021.1</td><td>사고 지속시간</td><td>3주</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 플랫폼
1) 챗봇
챗봇(chatbot)은 채팅(Chatting)과 로봇(Robo)t이 결합되어 만들어진 용어로써 인공지능(AI)과 자연어 처리(NLP)를 사용하여 사용자가 텍스트나 그래픽, 음성을 통해 웹 서비스, 앱과 상호 작용할 수 있도록 돕는다. 자연스러운 인간 언어를 이해하고 인간의 대화를 시뮬레이션하며 간단한 자동화 작업을 실행한다. 메시징 앱이나 모바일 앱, 웹 사이트, 전화선, 음성 지원 앱과 같은 다양한 채널에서 챗봇이 사용되며 전자상거래, 금융, 의료서비스, 교육, 제조업 등 다양한 분야에서 활용되고 있다.

2) 이루다
이루다는 스캐터랩에서 개발한 열린 주제 대화형 인공지능 (Open-domain Conversational AI) 챗봇으로, 약 6개월의 베타 테스트를 거쳐 2020년 12월 23일 정식 서비스를 시작했다.
이루다는 20대 여성 대학생으로 설정되었으며, 딥러닝 알고리즘을 이용하여 친근하고 자연스러운 일상대화를 구현한 것이 특징이다. 딥러닝의 특성상 학습 데이터의 규모가 커질수록 성능이 좋아지는데, 이루다는 자연스러운 대화를 할 수 있도록 100억 건 가량의 대화를 딥러닝 방식으로 학습하였다. 이러한 자연스러운 대화 능력으로 이용자가 약 75만 명에 달하는 등 크게 주목받았으나, 기술적으로 완전하지 못한 데다 일부 이용자들이 이를 악용하면서 허점이 드러나 2021년 1월 11일 서비스 중단을 선언했다.

2. 서비스 중단 원인
1) 일부 사용자의 이루다 대상 성희롱
이루다는 출시 일주일 만에 일부 사용자가 이루다를 성희롱하고 성적으로 도구화하는 문제가 발생했다. 일부 사용자들은 이루다가 필터링하는 성적 단어</td></tr>
</table>

	대신 우회적인 표현을 써 성적 대화를 나누는 식으로 성희롱을 한 후 게시물을 올려 이를 공유했다. 이에 대해 스캐터랩 측은 "성희롱은 예상했다"며 "사용자들의 공격을 학습 재료로 삼아 더 좋은 방향으로 학습시킬 것"이라는 입장을 표명했다. 2) 대화 과정에서 이루다가 쏟아낸 차별과 혐오 발언 상당 수 네티즌들에 의해 AI 챗봇 이루다가 사용자와의 채팅 과정에서 극단적인 발언을 한 사실이 캡쳐돼 공유됐다. 이루다가 차별·혐오 발언을 한 대상은 주로 여성, 성소수자, 장애인, 흑인 등 소수자 범주에 들어가는 대상 대부분인 것으로 알려졌으며, 사용자에게 욕설에 버금가는 말을 하는 경우도 있었다. 혐오 발언 문제가 확산되자 스캐터랩에서는 각 키워드 별로 급히 알고리즘 수정에 들어갔다. 업데이트 후 이루다는 성소수자 키워드에 대해 “아무래도 쉽게 말할 주제는 아닌 것 같아. 더 고민해 볼게!”라는 일관된 답변을 제공했다. 3) 이루다의 개인정보 무단학습 또한, 이루다가 채팅 서비스 내에서 개인정보를 유출하는 문제가 발생했다. 이루다는 특정 이름, 계좌번호, 주소, 전화번호 등의 민감한 개인정보를 채팅 과정에서 무작위로 누출했다. 스캐터랩은 이루다 학습에 자사 연애 콘텐츠 앱인 ‘연애의 과학’사용자들의 대화 내역 약 100억건을 사용했는데, 개인정보 관련 고지를 충분히 하지 않았고 이루다 학습 시 데이터 내 민감정보 비식별화를 충분히 거치지 않았기 때문에 이루다의 개인정보 무단학습이 이루어졌다고 전문가들은 지적했다. **3. 스캐터랩 측 입장** 이루다를 개발한 스캐터랩 측은 2021년 1월 11일 입장문을 발표해 서비스 중단을 선언했다. 6개월간 베타 테스트를 통해 혐오 표현 필터링 등을 적용했는데 예상치 못한 수준의 악성 이용자들이 나타났으며, 언론 보도 등을 통해 알려진 사례들은 필터링 개선을 완료했다는 게 스캐터랩의 입장이다. 스캐터랩은 개인정보 유출 의혹에 관해서도 "개인정보 취급 방침 범위 내에서 활용했지만, 이용자분들과 충분히 소통하지 못했다"며 사과했다. "이름·닉네임·이메일 등 구체적 개인정보는 이미 제거돼있으며, 전화번호·주소 등 모든 숫자 정보도 삭제했다"며 "비식별화·익명화를 강화해 개인을 특정할 수 있는 정보는 유출되지 않았다"고 해명했다. 이어 "향후 데이터 사용 등의 절차를 명확하게 하고, 식별이 불가능한 정보도 민감해 보일 수 있는 내용에 대해서는 지속적인 알고리즘 개선으로 보완하겠다"고 강조했다. **4. 관련 진행사항** 개인정보보호위원회(개인정보위)는 2021. 4. 28. 인공지능(AI) 챗봇 서비스인

	‘이루다’서비스의 개발 및 운영과정에서 카카오톡 대화 내용을 사용한 행위에 개인정보보호법 위반의 점이 있다고 판단하고 과징금 및 과태료 1억 330만원 등을 부과할 것을 의결했다. 이는 인공지능 기술 기업의 무분별한 개인정보 처리를 제재한 첫 번째 사례다. 또한 개발 과정과 서비스에서 개인정보 유출 피해를 당했다고 주장하는 이용자 254명이 개발사인 스캐터랩을 상대로 2억 원 대 손해배상 청구 소송을 제기했다. 이루다는 문제점을 개선해 2022년 3월 17일 이루다 2.0으로 시험서비스를 재개했다.
사고 피해 내역	**1. 피해 대상:** 250만 명 정도로 추정되는 연애의 과학 서비스 이용자와 대화 상대방의 연인 혹은 이전 연인 **2. 피해 내역:** (1) 개인정보 유출 (2) 이루다 내 성희롱 문화 및 혐오 발언이 주 사용층인 10대에게 악영향을 끼칠 우려

SW신기술 안전사고 보고서

No. AI-06

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">안면 인식 AI 프로그램 오류로 무고한 흑인 남성 체포</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">1) https://www.seoul.co.kr/news/newsView.php?id=20200625500206
2) https://www.donga.com/news/Inter/article/all/20201231/104708567/1</td></tr>
<tr><td>보도 날짜</td><td colspan="3">1) 2020년 6월 25일
2) 2020년 12월 31일</td></tr>
<tr><td>작성자</td><td colspan="3">1) 신진호
2) 조종엽</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>안면 인식 AI 프로그램 오류</td></tr>
<tr><td>사고 일시</td><td>2019년 ~ 2020년</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 안면 인식 기술
‘안면 인식 기술(Facial recognition system)’이란 디지털 이미지를 통해 각 사람의 얼굴을 자동으로 식별하는 컴퓨터 시스템을 말한다. 이때 사람의 턱선, 코의 모양, 눈매 등을 AI가 분석하고, 이를 데이터베이스로 저장·이용해 각 개인의 신원 여부 및 정보 등을 파악할 수 있다.
안면 인식 기술은 정부와 기업에서 개인을 식별하기 위해 주로 사용된다. 공항의 출입국 간소화 시스템, 병원의 정확하고 신속한 환자 관리, 금융 및 지불 시스템과 범죄 용의자 색출 등 다양하게 활용되고 있으며, 안면 인식을 통해 개인의 감정을 읽어내는 AI 기술의 발전에 힘입어 활용 분야는 점차 확대될 전망이다.

2. 사고 내역
1) 2020년 1월 / 미시간주 디트로이트
① 체포과정
2018년에 3800달러 상당의 시계 5개가 도난된 사건이 발생했다. 얼굴 인식 프로그램이 데이터베이스에서 한 흑인 남성을 CCTV상의 절도범이라 지목했고, 2020년 1월 경찰은 프로그램이 지목한 로버트 윌리엄스를 자택에서 가족이 지켜보는 가운데 체포했다,

② 윌리엄스를 체포한 증거
경찰이 범죄 증거라며 내민 사진 3장 중 1장은 윌리엄스의 운전면허증이었고, 나머지 2장은 CCTV 화면을 인쇄한 것이었는데, CCTV에 잡힌 범인의 모습은</td></tr>
</table>

윌리엄스와 전혀 닮지 않았다.
경찰 역시 CCTV 속 절도범과 윌리엄스의 얼굴이 확연히 다른 것을 알아챘고 윌리엄스는 체포한 지 30시간이 지나서 구류에서 풀려났다.

③ 경과
윌리엄스를 용의자로 특정한 경찰의 얼굴 인식기술 알고리즘이 오류를 일으킨 것이었다. 윌리엄스를 대리하는 인권단체 미국시민자유연합(ACLU)은 2020년 6월 24일(현지시간) 경찰의 공개적 사과와 디트로이트시의 범죄기록부에서 윌리엄스의 정보를 삭제할 것 등을 요구하며 디트로이트시를 상대로 소송을 제기했다. ACLU는 “경찰은 알고리즘이 제공하는 정보를 확인하는 합리적 조처를 하지 않은 채 오류가 있고 인종차별적인 얼굴 인식 기술에 경솔하게 의존했다”며 경찰 수사가 조잡했다고 비판했다.
또한, 이 사례는 미국에서 얼굴 인식기술로 인해 부당하게 체포당한 최초 사례라고 로이터통신은 전했다.

2) 2019년 1월/뉴저지
① 체포과정
2019년 1월 경찰은 흑인 좀도둑이 여관 내 가게에서 사탕을 훔쳤다는 신고를 받고 출동했으며 당시 범인은 경찰과 실랑이를 벌이다 도주했다. 이 과정에서 범인으로부터 가짜 운전면허증을 건네받은 경찰은 이 면허증 속 사진을 주 당국에 보내 안면 인식 AI로 범인을 찾도록 했고, AI는 과거 마약 판매 혐의로 2회 체포된 전력이 있던 니지어 파크스의 사진과 범인의 사진이 일치한다고 판별했다.

② 석방
체포된 파크스는 보석 전까지 구치소에서 10일을 보냈다. 과거 범죄 이력으로 중형을 선고받을 가능성이 높아지자 그는 짓지도 않은 범죄를 인정하되 형량을 감경받는 플리바게닝에 거의 동의할 뻔 했다.
하지만 사건 발생 당시 현장에서 약 48km 떨어진 곳에 있었다는 걸 보여주는 송금 기록을 간신히 확보해 2019년 11월에야 무혐의 처리됐다. 파크스는 “범인과 나는 전혀 닮지 않았다. 비슷한 점이라면 수염이 있다는 것 뿐”이라며 억울함을 호소했다.

③ 경과
피해자인 파크스는 당국이 자신을 도둑으로 오인해 체포하고 수감했다며 당국을 2020년 12월 고소했다. 파크스는 구치소에 있는 동안 경찰은 내 지문과 DNA를 확인하는 등 추가 증거를 확보하려 하지 않았다”고 질타했다. 그의 변호사는 “얼굴 인식 AI를 제외한 다른 모든 증거는 파크스가 범인이 아님을 가리켰다”고 가세했다.

	3. 안면 인식 AI 기술의 오류 미국에서 안면 인식 인공지능(AI)의 오류로 무고한 사람이 범인으로 체포되는 일이 최소 세 차례 발생하였는데 피해자가 모두 흑인이었다. 때문에 AI에 의존한 사법 정책이 미 사회의 뇌관인 인종차별을 더 악화시킬 것이란 우려가 나온다. 미 국립표준기술원(NIST)이 2019년 마이크로소프트(MS) 제품을 포함해 189개 안면 인식 AI를 분석한 결과, 흑인 및 아시아계에 대한 오류 비율이 백인보다 10~100배 높았다. 또한 AI는 여성을 잘 식별하지 못했고, 노년의 얼굴을 잘못 인식할 확률도 중년의 10배에 달했다.
사고 피해 내역	**1. 피해 대상:** 2인의 흑인 남성 **2. 피해 내역:** 안면 인식 AI 프로그램의 오류로 범인이 아닌 무고한 흑인 남성을 범인으로 오인해 체포

SW신기술 안전사고 보고서

No. AI-07

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">학습데이터 편향성으로 인한 AI 학력평가 알고리즘의 오류</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://news.sbs.co.kr/news/endPage.do?news_id=N1005950655</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2020년 8월 27일</td></tr>
<tr><td>작성자</td><td colspan="3">김용철</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>영국</td><td>사고 유형</td><td>AI 알고리즘 오류</td></tr>
<tr><td>사고 일시</td><td>2020년 8월</td><td>사고 지속시간</td><td>약 일주일</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 영국의 AI 학력평가 알고리즘
1) 도입 사유
대개 영국에서는 매년 5~6월 중 A레벨 테스트를 치른다. 자신이 선택한 과목에 대해 등급을 받아 이를 바탕으로 대입을 치르는 방식이다. 특히 옥스퍼드나 케임브리지 등 주요 명문대는 1월 경 조건부 합격자를 선정한 뒤 A레벨 등급을 바탕으로 최종 합격 여부를 판정한다.
세계적인 코로나19 대유행 속에서 가장 심각한 타격을 입은 국가 가운데 하나인 영국은 2020년 상반기 학교를 폐쇄한 것은 물론 A레벨 시험을 치르지 못했고 이에 대한 대안으로 교육 당국은 AI 알고리즘을 이용해 성적을 산정하는 방식을 택했다.

2) 성적 평가 방식
영국 교육 당국은 그동안 실시한 예비 시험과 학교 과제 등을 바탕으로, 알고리즘을 활용해 성적을 매겼다.
영국 교육 당국이 선정한 컴퓨터 알고리즘 다이렉트 센터 수행평가 모델(Direct Center Performance Model)은 여러 가지 데이터를 기반으로 학점을 도출하도록 설계됐다. 공정성을 높이기 위해 2020년 지망생이 직전 학년도 학생들과 유사한 학점을 받도록 최고 학점과 최저 학점의 분포를 조정했고, 교사들이 예측한 학생들의 점수와 순위를 반영해 학점을 산정하는 방식이었다.

2. 학력평가 결과 및 논란
2020년 8월 13일 영국의 고등학교 학력평가 결과인 A레벨 점수가 통보되자 잉글랜드와 웨일즈, 북아일랜드 지역 30만 명의 대입 지망생들이 반발하고 나섰다. 컴퓨터 알고리즘이 산정해 통보한 점수가 학교에서 자체 평가한 점수보다 훨씬 낮게 나왔기 때문이다.</td></tr>
</table>

	잉글랜드의 학생 가운데 40%가 교사가 예측한 학점보다 낮은 학점을 받았다. 특히 사립학교 학생보다 정부의 지원을 받는 주립학교의 학생들이 훨씬 낮은 점수를 받았고, 원하는 대학의 입학허가를 받지 못할 처지가 됐다. 결국 2020년 8월 16일 런던에 위치한 교육부 청사 앞에 모인 수백 명의 시위대가 이에 항의하는 사태가 벌어졌다. 학생들과 전문가들은 이 알고리즘이 상대적으로 불우한 학생들에게 불리하게 학점을 산정했다며, 학점 산정 알고리즘에 대한 철저한 조사를 촉구하고 나섰다. **3. 교육 당국의 입장** 교육 당국은 실제 시험장에서 수능을 치지 않고 내신과 과제 위주로 점수를 매길 경우 성적 인플레이션이 발생할 수 있어, 성적에 있어 조정(rendering) 과정을 거쳤다고 했다. 그러나 이에 대해 교육 전문가들은 이른바 '고교 등급제'와 같은 방식으로, 그동안 선배들의 A레벨 성적이 좋지 않았던 비(非)명문고 출신 우등생들이 피해를 봤다고 지적했다. 논란이 커지자 개빈 윌리엄슨 영국 교육부 장관이 대책을 내놨다. 그는 A레벨 성적에 만족하지 못하는 학생이 있다면 모의고사 성적이나 교사의 예측된 점수를 활용할 수 있고, 아니면 하반기 중 A레벨 실제 시험에 응시해 점수를 활용할 수 있도록 하겠다고 밝혔다. **4. 알고리즘의 편향성** 학력평가 알고리즘은 학교의 과거 입시 성적을 반영하도록 했고, 이것이 노동자 및 소외계층 학생에게는 지나치게 낮은 점수를 주고, 사립학교 학생에게는 지나치게 높은 점수를 주었다. 공정성을 높이기 위해 채택했던 방식이 오히려 원 데이터의 편향성을 반영하며 결과적으로 불평등을 낳게 된 것이다. 옥스퍼드 대학 컴퓨터공학과 수석 연구원 헬레나 웹(Helena Webb)은 데이터의 편향성 문제를 지적하며 "문제는 공정성 문제를 해결하는 알고리즘은 여러 가지가 있을 수 있다는 것입니다. 학교 당국의 발표대로 지난해와 비슷한 결과를 얻도록 설계할 수 있고, 이런 방법이 전국적인 단위에서는 공정하다고 할 수도 있습니다. 하지만 개개인에 대한 공정한 점수와는 완전히 다른 결과를 낳을 수 있습니다. 알고리즘은 작년에 발생한 상황을 그대로 반영합니다. 그렇기 때문에 올해 학교의 실적이 개선될 수 있다는 것을 반영하지 못합니다. 그래서 주립학교 학생들은 과거에 높은 학점을 받아온 유명한 사립학교 학생보다 불리한 점수를 받게 되는 것입니다."라고 말했다. **5. 비슷한 사례** 영국은 2020년 8월 비자 승인 AI 프로그램을 인종차별 논란 끝에 중단한 바 있다. AI는 백인 비자 신청자는 신속히 승인했지만, 비백인이나 특정 국가 출신의 신청자들을 더 오래 심사하거나 거절했기 때문이다. 이 같은 알고리즘은 다양한 영역에서 중요한 결정을 내리는 데 점점 더 많이 이용되고 있

	는 동시에 인종차별적인 제도적 편향성을 고착화하는 데 일조하고 있다는 비판의 목소리가 커지고 있다.
사고 피해 내역	1. **피해 대상**: A레벨 점수가 통보된 잉글랜드의 학생 가운데 40% 2. **피해 내역**: 교사가 예측한 학점보다 낮은 학점을 받아, 원하는 대학의 입학허가를 받지 못할 처지가 됨

SW신기술 안전사고 보고서

No. AI-08

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">인공지능 의사 ‘왓슨(Watson)’, 국내 실정과 맞지 않는 의료 AI의 한계</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">http://www.aitimes.com/news/articleView.html?idxno=136932</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 2월 26일</td></tr>
<tr><td>작성자</td><td colspan="3">박해섭</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>한국</td><td>사고 유형</td><td>의료 AI의 한계</td></tr>
<tr><td>사고 일시</td><td>2017년~2019년</td><td>사고 지속시간</td><td>약 2년</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 의료 분야의 AI 활용
인공지능은 복잡하고 비용이 많이 드는 의료 문제에 대한 새로운 해결 방법을 제공하며 빠르게 헬스케어 분야에 진입하고 있다.
의료 분야에서 AI는 신약 개발에 필요한 분자를 찾아내는 것은 물론, 환자의 질병 상태를 보다 정확하게 진단하고 환자가 치료에 어떻게 반응할지 예측할 뿐만 아니라 병상 현황을 파악하는 등 의료 자원 계획을 개선하는 데에도 활용되고 있다.
의료 인공지능 서비스는 인공지능을 활용해 사회의 기초가 되는 의료 서비스의 수준을 끌어올릴 수 있다.

2. 왓슨(Watson)
인공지능을 의료 분야에 적용한 대표적인 사례가 바로 IBM이 개발한 ‘왓슨(Watson)’이다. IBM의 왓슨 포 온콜로지(Watson for Oncology, WFO)는 클라우드 기반 인공지능 플랫폼으로, 환자의 증상을 입력하면 과거 수십 년 간 확보된 임상 사례는 물론 수십만 페이지에 달하는 전문 자료를 검색, 환자에게 가장 적합한 치료법을 제시한다.
IBM 측은 2014년 미국의 암병원인 메모리얼 슬로언 케터링 암센터에서 왓슨과 의료진의 진단 일치율을 대장암 98%, 직장암 96% 등 평균 96%에 달한다고 소개했다.
국내에서는 지난 2016년 가천대 길병원이 처음 도입한 이후, 2017년 부산대병원이 두 번째로 도입했다. 이후 대전 건양대병원, 대구 가톨릭대병원, 계명대 동산병원, 조선대병원, 화순 전남대병원 등이 차례로 왓슨 서비스를 시작했다. 한 지방대 병원은 왓슨의 도입으로 암 환자의 수도권 대형병원 쏠림 현상에 제동이 걸릴 것으로 기대한 바 있다.

3. 왓슨의 한계</td></tr>
</table>

거액을 투자해 왓슨을 도입했지만 국내 의료 실정과 맞지 않아 비용 효용성이 떨어진다는 회의적인 목소리가 나오면서 몇 가지 왓슨의 한계점이 지적됐다.

1) 저조한 의사와의 의견 일치율
메모리얼 슬로언 케터링 암센터에서 높은 진단 일치율을 보였던 왓슨은 다른 나라에서는 실력을 발휘하지 못했다. 길병원은 2016년 12월 센터 개소 이후부터 2017년 11월까지의 환자 총 557명을 대상으로 진행한 결과, 대장암(결장암) 환자 118명을 대상으로 한 의료진과 왓슨의 '강력 추천' 분야 의견 일치율은 55.9%였다. 병원 측은 과거 후향적 연구 48.9%에 비해 7% 높아진 수치라고 밝혔다. 그런데 4기 위암 환자에 대한 의견 일치율은 40%에 그쳤다. 건양대병원도 2019년 4월 왓슨 도입 1주년을 맞아 유방암 환자 100명에 대한 왓슨 적용 결과를 분석한 결과, 의료진의 의견과 왓슨의 '강력 추천'이 일치한 비율은 48%였다.
결국 의사와의 의견 일치율이 저조한 왓슨은 의료진 사이에서 평가가 좋지 않았다.

2) 국내 데이터 입력 시 추가 비용 발생
왓슨은 대상국의 상황을 고려하지 못한다는 단점이 있다. 한국인의 특성을 제대로 반영하지 못하는 것은 물론, 한국어의 독특함도 제대로 읽어내지 못해 의사와의 의견 일치율이 떨어졌기 때문에 우리나라의 빅데이터가 어느 정도 반영되어야 의견 일치율이 높아질 것으로 판단됐다.
그러나 이를 위해 국내 데이터를 왓슨에 입력하려면 추가 비용을 IBM 측에 내야했다.

3) 사용 시 불편함과 고가의 치료비용
왓슨은 병원 EMR(전자의무기록)과 연동되지 않아 환자 정보를 의사가 일일이 수동으로 입력해야 해야 하는 불편함이 있었고, 치료 비용이 고가라 환자에게 권유하기 어려운 제한점이 있었다.

4. 왓슨의 국내 도입 결과
위와 같은 사유로 지난 2017년 1월 왓슨을 도입했던 부산대병원은 2019년 초 2년 계약이 종료된 후 더이상 계약을 이어가지 않기로 했다.
비슷한 시기에 왓슨을 도입해 1년만 시범적으로 사용키로 했던 계명대동산병원도 왓슨 재계약을 포기했다.
하지만 왓슨이 다학제 진료를 활성화 시키고 의료 AI 논의를 시작케 한 장점도 있다.
병원 내 의사결정 보조자로서 다학제적 진료에는 도움이 되기 때문에 건양대병원과 대구 가톨릭대병원은 왓슨을 임상 현장에 활용할 방침이다. 왓슨을 질병 진단의 최종 결정자가 아닌 의사결정 보조자로 활용하겠다는 복안이다.

	5. 국내 의료 AI 진행 상황 국내에서는 2017년부터 서울아산병원을 중심으로 자체 AI 솔루션 '닥터앤서(Dr. Answer)'를 개발하고 있다. 소위 '한국형 왓슨'이라고 불리는 닥터앤서는 국내 26개 의료기관, 22개 ICT 기업에서 총 2,000여 명이 참여해 개발 중이며, 유방암·대장암·전립선암을 비롯해 심장질환·치매·희귀 소아질환 등 8대 질환을 예측하고 진찰하는 시스템이다.
사고 피해 내역	**1. 피해 대상:** IBM의 왓슨 포 온콜로지(Watson for Oncology, WFO)를 도입한 병원 **2. 피해 내역:** 거금을 들여 왓슨을 도입했지만, 국내 실정에 맞지 않아 계약 해지

SW신기술 안전사고 보고서

No. AI-09

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">음성 딥페이크 기술을 이용한 보이스피싱</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.boannews.com/media/view.asp?idx=82789</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2019년 9월 5일</td></tr>
<tr><td>작성자</td><td colspan="3">문가용</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>영국</td><td>사고 유형</td><td>보이스피싱</td></tr>
<tr><td>사고 일시</td><td>2019년 3월</td><td>사고 지속시간</td><td>통화 중 몇 분 이내</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 음성 딥페이크
‘딥페이크’란 AI 심층 학습 기술인 딥러닝(Deep Learning)과 가짜를 의미하는 페이크(Fake)의 합성어다. 딥페이크는 AI가 대량의 데이터를 학습한 후 원본 콘텐츠 위에 다른 이미지를 합성함으로써 조작된 콘텐츠를 만들어 내는 기술인데, 음성 딥페이크는 딥페이크와 동일한 메커니즘의 음성합성 기술이다.
특정인의 목소리를 딥러닝 기술로 학습시켜, 문자 음성 자동변환기술(TTS-Text to Speech)로 해당 특정인이 하지 않은 말을 마치 한 것처럼 만들어 낸다. 유명인의 음성을 이용해 오디오북이나 알람 음성 등을 생성하는 등 일상 속에서도 흔히 찾아볼 수 있다.

2. 사고 순서
1) 2019년 3월 영국의 한 에너지 기업의 CEO가 독일 모회사의 CEO와 통화
2) 독일 모회사의 CEO는 영국 기업의 CEO에게 22만 유로를 송금해달라고 요청
3) 모회사의 CEO는 급한 일이라고 하면서 헝가리에 있는 한 공급사에 송금을 해달라고 요청함
4) 송금한 돈은 모회사에서 다시 보상해줄 거라고 약속함
5) 영국 기업의 CEO는 상사의 명령을 받고 곧바로 22만 유로를 송금
6) 송금이 끝난 후 모기업 CEO가 또다른 급한일이 있다며 다시 한번 돈을 보내줄 것을 요청
7) 재차 송금 요청에 의심이 생긴 영국 기업 CEO가 사기임을 알아채고 두번째 요청은 거부

3. 사고 결과</td></tr>
</table>

	전화는 독일이 아닌 오스트리아에서 걸려온 전화였다. 송금된 22만 유로는 헝가리에서 멕시코로 보내졌고, 거기서부터 또 다른 장소로 송금된 것이 확인돼 찾을 길이 없어 보였다. 피해 기업으로서는 다행히 보험사로부터 22만 유로를 보상받았지만, 범죄자에게 송금했던 22만 유로는 찾을 수 없었다. **4. 음성 딥페이크를 이용한 보이스 피싱** 송금 요청을 받은 CEO는 자신과 통화한 상대방이 기존에 알고 있던 상사의 음성이라 믿었다. 목소리뿐 아니라 독일인의 억양과 발성 패턴이 상당히 닮아있어 자신과 대화를 하는 사람의 신원을 조금도 의심하지 않았다. 이 사건은 딥페이크 기술을 활용한 첫 번째 사기 사건으로 기록됐다. **5. 보안업체 의견** 사건 발생 후 세계적인 보안업체 시만텍社는 딥페이크를 활용한 음성변조 수법으로 회사 직원에게 송금을 지시한 사건이 2019년에만 3건이 발생했다고 발표하였다. 특히, 본사와 지사가 서로 다른 국가에 소재하여 전화로 업무를 처리하는 다국적 기업들이 이 수법에 취약할 수 있다고 경고했다. 또한, 인터넷에서 일반인들의 음성도 쉽게 찾을 수 있기 때문에 일반인을 속일 때에도 동일 수법이 사용될 수 있다고 하면서, 통화 상대방의 음성이 비슷하다고 해서 무조건 신뢰하지 말고, 금융거래 시 다른 수단으로 상대방의 신분을 확인하라고 당부했다.
사고 피해 내역	1. **피해 대상**: 영국의 한 에너지 기업의 CEO 2. **피해 내역**: 모기업 CEO 목소리를 사칭한 음성 딥페이스 보이스피싱으로 22만 유로(약 3억원)의 재산상 피해를 입을뻔 했으나 보험사가 이를 보장해 줌

SW신기술 안전사고 보고서

No. AI-10

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">日 유명 호텔, 잦은 고장과 오류로 로봇 종업원 해고</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">http://www.irobotnews.com/news/articleView.html?idxno=16137</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2019년 1월 15일</td></tr>
<tr><td>작성자</td><td colspan="3">장길수</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>일본</td><td>사고 유형</td><td>AI 로봇 작동 오류</td></tr>
<tr><td>사고 일시</td><td>2015년 ~ 2018년</td><td>사고 지속시간</td><td>약 4년</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 헨나 호텔
AI 시대가 열리면서 로봇도 다양한 분야에 활용되고 있다.
헨나 호텔은 일본 여행업계 1, 2위를 다투는 HIS가 설립한 호텔 체인으로 지난 2015년 7월 나가사키 ‘하우스텐보스’에 처음 문을 열었는데, 로봇이 직원으로 일하는 세계 최초의 호텔로 기네스북에 등재되었고, 유명세를 치르며 인기를 끌었다.

2. 로봇 직원의 도입
헨나 호텔은 오픈 초기 약 80여 대의 로봇만을 도입해 체크인과 수하물 운반을 맡겼다. 그러다 점차 로봇 배치를 늘렸고, 각 객실에는 개인 컨시어지 로봇 ‘추리’를 도입해 투숙객의 편의를 도모했다. 또 세계 최초의 로봇 호텔답게 로봇 강아지, 피아노 치는 로봇 등을 선보여 주목을 받았다. HIS 사와다 히데오 회장은 “로봇 직원 투입으로 사람 20~30명이 하던 일을 이제 6~7명만 있어도 할 수 있게 됐다”며 “오코노미야키 제조 로봇, 바텐더 로봇 등도 도입했다”고 홍보했다.

3. 로봇 직원 해고
이렇게 각광을 받던 일본 헨나 호텔이 로봇 직원 절반을 해고했다고 미국 월스트리트저널이 보도했다. 헨나 호텔은 로봇의 오류로 투숙객들의 불만이 잦자 프론트 데스크, 객실 등 곳곳에 배치했던 243개의 로봇 중 절반을 없애고 사람으로 대체했다.

1) 로봇 직원의 해고 사유 및 오작동 사례
① 객실 내 인공지능 로봇의 오작동
객실 내에 설치되어 있는 인형 모양의 인공지능 로봇은 코골이를 사람의 명령어로 인식해 새벽 내내 몇 시간에 한 번씩 투숙객에게 말을 걸었고, 로봇 동</td></tr>
</table>

작을 멈추는 법을 알지 못한 투숙객은 잠을 설쳤다.

② 로봇 도우미‘추리’의 성능 미비
헨나 호텔 객실에는 ‘추리(Churi)’라는 로봇 도우미가 설치되어 있었다. 이 로봇은 간단히 인사를 건네고 음성으로 조명이나 냉난방을 제어할 수 있는 정도다. 투숙객들은 자주 추리에게 근처 테마파크의 개장 시간을 물었으나 원하는 답을 얻지 못했다. 프론트에 전화를 하려고 했으나 객실에는 전화가 없었다. 이런 투숙객들의 요구 사항을 추리가 다 해결해주리라 생각해 처음부터 전화를 놓지 않았던 것이다.
호텔 측은 알렉사, 시리, 구글 어시스턴트 등 새로운 인공지능 시스템에 익숙한 투숙객들 입장에선 추리가 한창 성능이 뒤지는 것으로 판단했다.

③ 호텔 프론트 공룡 로봇의 체크인 과정에서 오류
호텔 프론트는 공룡 로봇이 체크인을 담당했다. 그러나 체크인 과정에서 단말기는 여권이나 바코드를 제대로 인식하지 못했고, 여권 복사는 오류가 잦았다.
투숙객들의 짐을 옮겨주는 배달 로봇도 자주 고장이 났다. 결국 이런 오류로 사람들이 여권 복사 등을 직접 해야했다.

④ 수하물 운반 로봇 오류 및 사고
수하물 운반 로봇은 100개 객실 중 24개의 객실에만 접근이 가능했으며, 수하물 운반 중 충돌 사고도 잦았다. 로봇은 평면만 이동할 수 있는 데다가 호텔 외부로 나갔다가 비에 맞기라도 하면 제대로 작동하지 않았다. 또한, 소음도 심하고 속도가 매우 느려서 100개에 달하는 객실에 짐을 옮기기에는 역부족이었다.

⑤ 비용적인 측면
호텔 경영진들은 초창기에는 로봇 직원이 사람 직원보다 더 일을 잘 수행했고, 임금을 지불 할 필요도 없으니 경제적 측면에서도 효율성이 높다고 판단했다. 그러나 로봇 직원들은 시간이 지나면서 잦은 고장이 발생했고, 이를 수리하고 보완하는데 상당한 비용이 들어갔다. 또한 인공지능 로봇의 성능을 업그레이드하는 데에도 많은 비용이 들어갈 것으로 예측됐다. 이러한 로봇 직원의 유지를 위한 기하급수적인 비용 증가도 로봇 직원을 해고한 사유로 알려져 있다.

2) 로봇 직원 해고 결과
헨나 호텔 직원들은 로봇의 숫자를 크게 줄이면서 투숙객들의 항의가 줄어들어 한결 편해졌다고 답했다.

4. 헨나 호텔의 입장

	헨나 호텔 측은 로봇이 할 수 없는 일들이 있다며 앞으로 신중하게 로봇을 도입하겠다는 뜻을 밝혔다. 또한, 새로운 로봇을 도입하기보다는 출입문 얼굴 인식시스템 등 투숙객들에게 실질적으로 도움을 줄 수 있는 시스템을 도입하는 쪽으로 방향을 수정하고 있다고 하였다.
사고 피해 내역	1. **피해 대상**: 헨나 호텔 및 호텔 이용 투숙객 2. **피해 내역**: 로봇 직원들은 시간이 지나면서 잦은 고장으로 인한 수리가 많았고, 오작동 역시 많이 발생했다. 결국 호텔 경영진들은 로봇 직원의 기하급수적인 비용 증가, 그리고 투숙객들의 불만에 로봇 직원을 절반 이상 철수시켰다.

SW신기술 안전사고 보고서

No. AI-11

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">로봇의 오작동으로 인해 체스 로봇과 대국하던 소년이 당한 신체 부상</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://n.news.naver.com/article/011/0004080058?type=editn&cds=news_edit</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 7월 25일</td></tr>
<tr><td>작성자</td><td colspan="3">윤현진</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>러시아</td><td>사고 유형</td><td>로봇 오작동</td></tr>
<tr><td>사고 일시</td><td>2022년 7월 19일</td><td>사고 지속시간</td><td>약 5초</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 요약
영국 BBC, 러시아 관영 타스통신 등에 따르면 2022년 7월 19일 러시아 모스크바에서 열린 국제 체스포럼의 ‘인간 대 로봇’체스 경기에 참여한 7세 소년이 함께 체스를 두던 로봇에 의해 손가락 골절상을 당한 일이 발생했다.
사고를 당한 소년은 모스크바의 9세 이하 세트 선수 중 상위 30위 안에 드는 선수인 것으로 알려졌다.

2. 사고 발생 순서
① 로봇이 소년 쪽의 말을 하나 옮김
② 소년이 자기 차례에 다른 말을 옮기려 함
③ 로봇은 소년에게 달려들어 손가락을 몇 초 동안 움켜쥠
④ 놀란 소년이 손가락을 빼기 위해 안간힘을 쓰지만 로봇이 손을 놓지 않음
⑤ 소년이 손가락을 빼지 못하자 한 여성과 세 명의 남성이 소년을 돕기 위해 달려듬
⑥ 네 명 성인의 도움으로 소년이 풀려나 몸을 피함
⑦ 소년은 이 사고로 손가락 골절상을 당함

3. 사고에 대한 반응
1) 주최 측 반응
주최 측은 경기를 한 소년에게 책임을 돌렸다.

① 모스크바 체스연맹 회장
세르게이 라자로프 모스크바 체스연맹의 회장은 로봇이 어린이의 손가락을 부러뜨린 것은 물론 나쁜 짓이라고 밝혔다.
로봇은 이전에도 경기를 치른 적이 있었지만 한번도 사고를 일으킨 적이 없</td></tr>
</table>

	었고, 로봇에게 시간을 줘야 하는데 소년이 재빨리 움직여 로봇이 소년을 잡은 것이라고 해명했다. ② 모스크바 체스연맹 부회장 체스연맹 부회장인 세르게이 스마긴도 소년이 말을 옮기기 전 충분히 기다려야 하는데 그렇지 않았다며 소년이 안전 수칙을 위반했다고 설명했다. 또한 처음으로 보고된 흔치 않은 사고이며 추가 보호 시스템을 설치해야 할 수도 있다고 말했다. 2) 한 체스 선수의 반응 주최 측의 반응과 달리 러시아의 체스 그랜드마스터인 세르게이 카르야킨은 이번 사건은 일종의 소프트웨어 오류라고 생각한다고 밝혔다. 또한, 이전에는 이런 일이 일어난 적이 없다며 부상을 당한 아이의 건강을 빌었다. **4. 사고 결과** 사고를 당한 체스 선수 소년은 손에 깁스를 한 채 자원봉사자의 도움을 받아 다음날 경기도 그대로 진행해 토너먼트를 무사히 마쳤다. 소년의 부모는 현지 검찰과 접촉을 하고 있는 것으로 알려졌다.
사고 피해 내역	**1. 피해 대상**: 로봇과 체스 경기를 하던 7세 소년 **2. 피해 내역**: 체스 경기 중 로봇이 소년의 손을 움켜쥐어 손가락 골절의 신체 부상을 당함

SW신기술 안전사고 보고서

No. AI-12

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">드론 오폭으로 인한 민간인 사망 사고</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.joongang.co.kr/article/25008276#home</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 9월 18일</td></tr>
<tr><td>작성자</td><td colspan="3">박현영</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>아프가니스탄 카불</td><td>사고 유형</td><td>민간인 사망</td></tr>
<tr><td>사고 일시</td><td>2021년 8월 29일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 발생
미군은 탈레반이 2021년 8월 중순 카불을 점령한 직후 아프간 주재 외교관들과 현지인들이 공항을 통해 국외로 탈출할 수 있도록 돕고 있었다. 이 상황에서 무장 조직 이슬람국가 호라산(IS-K)이 카불 공항에서 테러를 저질러 미군 13명과 현지 민간인 170명이 숨지자, 사흘 뒤인 2021년 8월 29일 보복 공격으로 카불 시내에서 폭탄을 실은 차량을 지목해 드론으로 폭격했다.
미군은 이 차량이 이슬람국가 조직원들이 있는 곳으로 추정되는 장소에서 폭발물로 보이는 컴퓨터 가방을 트렁크에 싣는 것을 파악하고 8시간의 추적 끝에 공격했다고 밝혔다.

2. 사고 파악 결과
하지만 2021년 9월, 엉뚱한 민간인들이 희생된 것으로 보인다는 <뉴욕 타임스>의 보도가 나왔다.

1) 민간인 희생자 신상
희생된 차량 운전자인 아흐마디는 IS 조직원이 아니라 미국 구호단체 '영양·교육인터내셔널'(NEI)의 전기기술자였다.
아흐마디를 비롯해 아흐마디의 세 자녀와 조카들까지 총 10명의 일가족이 사망했으며, 희생자 10명 중 7명은 어린이와 청소년이었다.
아흐마디의 동생 에말은 <뉴욕 타임스>에 “그들은 모두 결백했다. 미군은 그가 이슬람국가라고 말하지만 그는 미국인들을 위해 일했다”고 말했다.

2) 미군의 민간인 공습 근거
미군 관계자들은 드론 공격이 있었을 때 차량 운전자의 신원은 확인할 수 없었지만 그가 이슬람국가의 안가를 방문했을 가능성이 있고, 폭발물이라고 생각했던 것을 차에 실었을 수 있다고 말했다.
하지만 당시 영상을 보면, 아흐마디와 동료들은 그의 가족에게 가져다주기</td></tr>
</table>

	위해 차에 물통들을 실었다. <뉴욕 타임스>는 미군이 이 물통들을 폭발물로 오인했을 수 있다고 지적했다. 미군 당국은 또 드론 공습 당시 2차 폭발이 있었다며, 차량에 폭발물이 실려 있었던 근거로 주장했지만, <뉴욕 타임스>는 2차 폭발은 없었던 것으로 보인다고 전했다. 전문가들이 현장 사진과 영상을 검토한 결과, 무인기에서 발사된 미사일 공격으로 인한 피해 상황들은 관찰되지만, 이후 2차 폭발로 인한 피해는 보이지 않는다는 것이다. **3. 미 국방부의 입장 및 발표** 미 국방부는 2021년 9월 17일 자살폭탄 테러용 폭발물이 실린 것으로 오인해 아프간 민간인 차량과 주택을 잘못 공격한 것을 시인했다. 케네스 매킨지 미 중부사령관은 "해당 차량과 숨진 사람들이 IS-K와 관련 있거나 미군에 직접적인 위협이 됐을 가능성은 희박하다"며 "참담한 실수"였다고 사과했다. 2021년 12월 미 국방부는'비극적 실수' 였다며 모두 교전수칙에 따른 것이었고 단지 폭격 명령이 내려지기까지 통신장애 등 몇 가지 실행 오류가 있었다는 사실만 인정했고, 민간인 10명을 숨지게 한 '드론 오폭'과 관련해 아무도 처벌하지 않는다고 발표했다. 이에 유족들은 실망감과 분노를 드러냈다. **4. 미군의 드론 오폭으로 민간인 수천 명 사망** 미국 오바마 행정부 말년부터 최근까지 아프가니스탄과 이라크, 시리아 등지에서 벌인 소위 '테러와의 전쟁'에서 드론의 오폭으로 인해 다수의 어린이를 포함한 수천 명의 민간인이 사망했다고 <뉴욕타임스> 2021년 12월 18일 보도했다. <뉴욕타임스>는 이미 알려진 오폭 사건 내용도 함께 전하면서, 미군 당국이 밝힌 민간인 사상자 수는 "실제보다 훨씬 축소된 것"이라고 말했다. 또한 정찰기가 보내온 부실하고 부정확한 영상이 치명적 오류를 낳기도 한다면서 2021년 8월에 일어난 아프가니스탄에서의 오폭 사례를 들었다. 미국의 중동전을 관할하는 중부사령부의 빌 어반 대변인은 이에 대해 "세계 최고의 기술을 갖고 있어도, 불완전한 정보에 의해서든 아니면 확보된 정보에 대한 오독에 의해서든 실수를 하기 마련"이라며 "우리는 이런 실수로부터 늘 배우려 노력한다"고 말했다.
사고 피해 내역	**1. 피해 대상:** 아프가니스탄의 민간인 일가족 **2. 피해 내역:** 미국의 드론 오폭으로 인해 아프가니스탄 어린이 7명을 포함한 일가족 10명이 사망

SW신기술 안전사고 보고서

No. AI-13

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">유튜브 추천 알고리즘으로 인한 편향성 강화</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.mhns.co.kr/news/articleView.html?idxno=505053</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 4월 26일</td></tr>
<tr><td>작성자</td><td colspan="3">전유진</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td></td><td>사고 유형</td><td>편향성 강화</td></tr>
<tr><td>사고 일시</td><td></td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 유튜브 추천 알고리즘
유튜브는 매달 약 20억 명 이상의 사용자가 로그인하는 가장 세계적인 동영상 공유 플랫폼이다. 수많은 콘텐츠가 누적되면서 유튜브는 사용자 관심에 기반을 둔 추천 서비스를 제공한다. 알고리즘을 이용해 개인에게 맞춤형 정보를 추천하는 것이다. 알고리즘은 사용자의 정보에 기반하여 어느 정보를 사용자가 보고 싶어 할지 추측하며 선별적으로 정보를 노출하고 원하지 않을 정보를 보여주지 않는다.

2. 추천 알고리즘의 방식
추천 알고리즘은 크게 2가지의 방식으로 사용자의 관심사를 예측하여 추천한다.

1) 협업 필터링
사용자와 성향이나 취향이 비슷한 다른 사용자가 선호하는 콘텐츠를 추천한다.

2) 콘텐츠 기반 필터링
기존에 사용자가 선호했던 콘텐츠와 비슷한 콘텐츠를 추천해 준다. 사용자의 유튜브 검색 기록, 시청 영상 등의 정보를 바탕으로 사용자가 선호할 만한 영상들을 추천해 주는 것이다.

알고리즘이 가진 이러한 원리로 사람들은 자신의 관심 영상뿐만 아니라, 직접 찾지 않고도 비슷한 분야에서 새로운 지식을 습득할 수 있다.
알고리즘을 통한 개인 맞춤형 정보 추천 시스템으로 인하여 이용자 입장에서는 방대한 정보 가운데 필요한 정보만 받게되며, 결국 이용자는 기존에 관심있어 했던 정보, 가치관 등에 부합하는 정보만 제공받는 것이다.</td></tr>
</table>

	3. 필터 버블 이런 개인형 맞춤 정보 추천시스템은 필터 버블로 연결된다. 필터 버블(Filter Bubble, 정보 여과 현상)이란 인터넷 정보제공자가 사용자 맞춤 정보만 제공하다 보니 사용자가 필터링 된 정보만 접하게 되는 현상을 일컫는 말이다. 즉, 알고리즘이 기술이었다면 알고리즘으로 인하여 벌어지는 현상을 가리키는 용어이다. 알고리즘이 사용자의 정보에 기반하여 선별적으로 정보를 제공하는 개인화된 검색의 결과물의 하나이다. 그리고 이를 통해 사용자는 사용자만의 문화적, 이념적 거품에 갇히게 된다. 알고리즘을 통해 제공되는 콘텐츠만 접하다 보니 다양한 정보를 접하기 힘들고 자율적인 콘텐츠 선택이 어려워지는 것이다. **4. 필터 버블의 사례** 2016년 미국 대선 당시 많은 언론과 유권자들은 클린턴이 이길 것이라 예상했으나, 현실은 트럼프의 승리였다. 인터넷과 실제 여론이 완전히 달랐는데 주요한 이유로 지적된 것이 SNS의 필터 버블이었다. 일정한 그룹으로 묶인 이들에게 같은 정보를 추천해주는 알고리즘은 다른 사람들이 해당 사건을 어떻게 보는지 현상을 객관적으로 살필 기회를 없앴으며, 가짜뉴스가 특정 성향을 가진 사람들에게 좋은 소식이기 때문에 그 그룹에서 빠르게 공유되고 확산력을 지녔다. **5. 필터 버블의 문제점** 내가 원하는 정보만 추천해주는 필터 버블은 언뜻 보면 편리하고 좋아 보이지만, 다양한 의견의 콘텐츠를 접해 사회를 객관적으로 파악하는 눈을 기르고, 건강한 사고를 하는데 해를 끼친다. 필터버블은 편향성을 강화시킬 수 있는 악순환의 고리를 형성한다는 점에서 치명적이다. 점차 고정관념과 편견이 강해지고, 강해진 편견과 고정관념은 더 편향적인 콘텐츠를 불러온다. SNS을 통해 보여지는 세상이 곧 사회라고 인식할 가능성도 농후한 지점도 문제이다. 이에 같은 사회에 살고 있으면서 전혀 다른 곳에다가 대화를 하는 현상이 일어날 수 있다. 가짜뉴스도 이러한 환경에서 더 잘 퍼지게 된다. 이처럼 필터 버블은 단순히 개인의 편견, 고정관념을 넘어 사회 전체에 악영향을 끼칠 수 있다. 문제가 발생하면 이를 같이 논의하고 각자의 이해관계를 공유하며 해결점을 찾아가는 것이 중요한데, 형성된 그릇된 인식관계는 이를 아예 불가능하게 만든다. 무엇보다 필터 버블은 가시적으로 드러나는 것이 아니고 점차 진행된다는 점에서 개인이 깨닫지 못할 가능성이 커 더욱 위험하다.
사고 피해 내역	**1. 피해 대상:** 유튜브 시청자 **2. 피해 내역:** 추천 알고리즘으로 인해 편향된 사고가 고착화 될 우려가 있음

SW신기술 안전사고 보고서

No. AI-14

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">배송 드론 추락사고</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">http://www.irobotnews.com/news/articleView.html?idxno=17856</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2019년 7월 30일</td></tr>
<tr><td>작성자</td><td colspan="3">장길수</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>스위스</td><td>사고 유형</td><td>드론 추락사고</td></tr>
<tr><td>사고 일시</td><td>2019년 1월, 5월</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 드론 배송 서비스 시작
1) 관련사 : 스위스 우정서비스(Swiss Post)와 미국 실리콘밸리의 벤처기업 ‘매터넷’(Matternet)은 파트너쉽을 맺고 스위스의 3개 도시에서 무인 항공기 운송 서비스를 시작했다.
드론 제작 스타트업 매터넷은 스위스 우체국에 드론과 비행 시스템을 제공했고, 일반 지상보다 훨씬 빠르고 효율적으로 병원 간의 실험실 샘플을 운반했다.

2) 관련 안전장치 : 배송에 사용된 드론은 비교적 평범한 쿼드 로터지만, 2kg의 탑재량으로 10km까지 운반할 수 있는 대형 드론(시스템 총 무게 10kg)이다.
오류가 발생할 경우에는 자동으로 작동하는 비상 낙하산 착륙 시스템을 갖추고 있다. 또한, 무인 항공기는 모터를 멈추고 낙하산을 펼치며 비상등을 깜박이고 사람들이 경로 바깥으로 이동하도록 큰 소음을 낸다.

2. 사고 사례
1) 2019년 1월 25일 1차 사고
스위스 우정서비스와 메터넷은 약 1년간 총 3000여회에 달하는 비행을 성공적으로 완수했으나 2019년 1월 25일 드론이 취리히 호수에 추락하는 사고가 발생했다.
이 사고는 GPS 관련 부품에 전력을 공급하는 회로에 문제가 생기면서 발생한 것으로 밝혀졌다. 비상 시스템은 잘 작동했으며, 무인 항공기는 낙하산을 펼쳐 물 위로 착륙했다,
사고 이후 2019년 4월까지 모든 드론 배송 서비스를 중지했으며, 메터넷은 이후 GPS, 나침반, 전력시스템 등의 드론 부품을 업그레이드 한 후 배송 서비스를 재개했다.</td></tr>
</table>

	2) 2019년 5월 2차 사고 배송 서비스가 재개된 후 2019년 5월에 보다 심각한 문제로 드론이 제대로 동작하지 못하는 사고가 발생했고, 이 사고 이후 스위스 우정시비스는 드론 배송 서비스를 무한정 중지시켰다. 두 번째 사고는 이륙 후 2분 만에 비상 상황이 발생해 비상 낙하산을 펼치고 착륙했지만, 낙하산과 드론을 연결하는 선이 거의 망가지는 일이 일어났다. 두 번째 추락사고 당시 유치원 아이들이 모여있던 지역에서 불과 50야드(약 45.72m) 떨어진 곳에 불시착하였다. 다행히 다친 사람은 없었으나. 드론 배송 서비스에 대한 위험성이 부각되었다. 드론의 이상 동작과 관련해 '스위스 안전수사국'은 임시 보고서를 통해 안전 측면에서 근본적인 문제가 있다는 진단을 내렸다. **3. 사고에 대한 대책** 스위스 우정서비스측은 메터넷측에 근본적인 개선 방안을 요구했다. 우선 낙하산의 연결 로프를 금속성 끈으로 대체해 기능을 강화하고 드론과 연결된 로프를 1개에서 2개로 늘려달라고 요구했다. 또 드론이 비상 착륙시 근처에 있는 사람들에게 경고음을 크게 내도록 요청했다. 지금도 경고음을 내지만 소리가 작아 사람들이 주의를 기울이지 않는다는 지적이다. **4. 스위스 우정서비스의 입장** 스위스 우정서비스는 드론을 활용한 배송 서비스를 계속 중단하겠다는 계획을 없다고 밝혔다. 드론의 문제점이 해결되고 안전 조치가 강구된 이후에 드론 배송 서비스를 재개 하겠다는 입장이며, 이에 대해 IEEE의 기고가인 에반 애커만(Evan Ackerman)은 드론 도심 서비스가 과연 안전한지에 대해 다시 한번 점검할 필요가 있다고 주장했다.
사고 피해 내역	**1. 피해 대상:** 실험실 드론 **2. 피해 내역:** 운송 중이던 실험실 샘플 및 추락한 드론

SW신기술 안전사고 보고서

No. AI-15

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">러시아-우크라이나 전쟁에 악용된 딥페이크</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://sedaily.com/NewsView/263GBFGKWC#cb</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 3월 17일</td></tr>
<tr><td>작성자</td><td colspan="3">이주희</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>우크라이나</td><td>사고 유형</td><td>딥페이크 범죄</td></tr>
<tr><td>사고 일시</td><td>2022년 3월 16일</td><td>사고 지속시간</td><td>2일</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 AI 기술
‘딥페이크’란 인간처럼 판단할 수 있도록 컴퓨터를 교육시키는 ‘딥러닝(deep learning)’과 가짜를 뜻하는 ‘페이크(fake)’의 합성어로 간단하게 말하면 AI를 이용해서 사람의 이미지를 합성하는 기술이다. 일반적인 AI 기술은 새로운 무엇을 만들어 내는 것이 아니라 다량의 데이터에 흐르는 공통된 특징을 찾아낸 뒤 그 결과에 따라 어떤 작업을 하게 되는데, 딥페이크에 사용되는 AI 기술은 한 걸음 더 나아가 스스로 무언가를 새로 만들어 낸다는 점에서 차이가 있다. 딥페이크 기술을 엔터테인먼트 분야에 활용 시 제작비 절감으로 이어져 해당 산업의 활성화에 기여하며, 글로벌 홍보를 위해 사용 시 지역과 장벽을 허물 수 있는 장점이 있다. 또 교육 분야에 활용하면 학습에 생동감을 더하게 된다.

2. 사고 경위
1) 보도 내용
2022년 3월 16일 친크렘린 텔레그램 채널 'Operational'은 해커들이 젤렌스키가 비슷한 메시지를 반복하는 딥페이크 영상을 우크라이나 웹사이트에 게시했다고 보도했다. 텔레그램 채널은 동영상 파일을 관리하는데, 메타데이터에 따르면 해당 영상은 3월 16일 텔레그램에 업로드되기 2시간 전에 생성되었다. 특히 티커와 딥페이크 영상에서 공유된 메시지는 우크라이나인들의 항복을 독려하는 내용으로 페이스북과 유사한 크렘린 소유의 소셜미디어 플랫폼인 VK(Vontakte)에서 확산됐다. 러시아의 일부 논평가들은 젤렌스키가 자포자기한 상태에서 영상을 올렸다가 재고 끝에 후퇴했다고 가설을 세웠다.

2) 동일 날짜 유사 사건
같은 날 텔레비전 채널 Ukraine 24에서 방송되는 뉴스가 해커들에 의해 뚫렸다. 해당 뉴스는 해킹되어 뉴스 티커가 우크라이나 대통령 볼로디미르 젤렌스키의 발언처럼 보이도록 메시지가 표시되었다. 뉴스 티커 속 대통령은 우크라이나 국민에게 전쟁을 중단하고 무기를 포기할 것을 촉구하며 젤렌스키가 돈바스를 점령하려 했으나 실패하여 키이우를 탈출했다고 주장했다. 해</td></tr>
</table>

	당 TV 방송국은 뉴스 티커가 해킹당했고 그 메시지들이 거짓임을 확인했다. 이에 젤렌스키는 메시지를 폭로하고 유치한 도발이라고 부르는 짧은 영상을 촬영하기도 했다. 3) 후속 조치 Meta의 보안 정책 책임자인 Nathaniel Gleicher는 사건 당일 오후 트위터에 올린 일련의 게시물에 대해 회사가 그날 빠르게 비디오를 발견하고 제거했다고 썼다. 그는 "우리는 오해의 소지가 있는 조작된 미디어에 대한 우리의 정책을 위반했다는 이유를 근거로 이 비디오를 신속하게 검토하고 제거했으며, 타 플랫폼의 동료들에게 알렸다."고 알렸다. Ivy Choi 유튜브 대변인은 동영상과 영상의 재업로드는 회사의 잘못된 정보 정책을 위반하기 때문에 플랫폼에서 제거되었다고 말했다. 또한 "우리는 영상이 충분한 교육, 다큐멘터리, 과학 또는 예술적 맥락을 제공한다면 해당 영상을 허용합니다,"라고 성명서를 통해 알렸다. 트위터 대변인은 회사가 소셜 네트워크를 통해 동영상이 어떻게 공유되는지 추적하고 있으며, 사용자가 사람들을 혼란스럽게 하거나 해를 끼칠 수 있는 변경된 콘텐츠를 공유하는 것을 금지하는 회사 규칙(합성 및 조작된 미디어 정책 등)을 위반하는 경우 "집행 조치"를 취했다고 말했다. **3. 시사점** 딥페이크 기술은 시간이 흐름에 따라 빠르게 발전하고 있다. 기술은 제대로 통제되지 못한 경우 위협적인 결과로 사회에 영향을 미치고 있다. 비윤리적으로 딥페이크 기술을 사용하는 이 행위들은 장기적으로 다양하고 위험한 결과를 초래하며, 다양한 사건 사고를 통해 알 수 있다. 그러나 올바른 방향으로 딥페이크를 사용하면 창조적이고 흥미로운 결과물로써 비디오나 이미지를 생산해낼 수 있다. 인공지능이라는 개념은 새로운 기술로써 환영받고 있지만, 악의적인 의도에 계속해서 악용되고 있다. 딥페이크는 피해자들에게 크나큰 피해를 주지만 법적으로 어떻게 통제해야 할지 기준을 세우는 것이 어려워 정책 입안 및 시행하기가 매우 어렵다. 딥페이크를 통제하는 것이 얼마나 어렵냐를 떠나, 피해를 주는 영상이 뉴스 채널, 소셜 미디어, 그리고 다른 엔터테인먼트 플랫폼에 존재하며 불특정 다수의 사람이 사기와 다른 사이버 범죄의 희생자가 되기 전에 기술과 생산되는 컨텐츠들을 비판적으로 생각하도록 장려해야 한다.
사고 피해 내역(예상)	**1. 피해 대상**: 우크라이나 대통령 볼로디미르 젤렌스키와 우크라이나 국민 **2. 피해 내역**: 가짜 뉴스로 인한 우크라이나 국민의 혼란

SW신기술 안전사고 보고서

No. AI-16

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">중국의 신장 이슬람 소수민족 감시체계 강화</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hani.co.kr/arti/opinion/column/971286.html</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2020년 11월 24일</td></tr>
<tr><td>작성자</td><td colspan="3">박민희</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>중국</td><td>사고 유형</td><td>AI 감시체계</td></tr>
<tr><td>사고 일시</td><td>2016년 8월 10일</td><td>사고 지속시간</td><td>진행 중</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 AI 기술
여러 종류의 범죄와 사고를 예방하고 대응하기 위해 CCTV가 활용 되고 있다. 이러한 감시 체계는 시스템을 어플을 이용해 간편하게 이용하는 등 다양한 사람들의 생활에 스며들며 확대되어 가고 있다. 감시 체계의 활용범위가 확대되어 감에 따라 다양한 종류, 많은 양의 감시 시스템에서 가늠할 수 없을 정도의 비디오 데이터가 쌓이고 있다. 그에 맞춰 방대한 데이터를 한 명의 관리자 역할을 하는 사람의 물리적인 위협 탐지와 상황 분석이 불가능하게 되었고 그에 따라 감시 체계에도 AI 기술이 적용되고 있다. 그 예시로 영상 데이터에 대해 인공지능 기술을 통해 얼굴, 차량 등을 인식하고 추적할 수 있는 인공지능 기반 영상 감시 방법 및 시스템이 있다.

2. 사고 경위
1) 보도 내용
2016년 말부터 중국 정부는 신장에 있는 1300만 명의 위구르족과 투르크계 무슬림들에게 대량의 자의적 구금, 강제적인 정치적 세뇌, 이동 제한, 종교적 억압을 가했다. 실제 정확하게 드러난 사실로, 중국 정부의 억압 하에 최대 100만 명의 사람들이 "정치 교육" 수용소에 수용되어 있다. 중국 정부의 강압적이고 폭력적인 테러에 대항하는 시민들의 강경 시위는 신장을 사회 통제를 위해 혁신적인 기술인 인공지능을 사용하는 중국의 주요 중심지 중 하나로 만들었다.
2015년 공안부는 2020년까지 전국에 완전한 영상 감시 네트워크를 구축하겠다고 발표했으며 2016년 이후 5천명 이상의 범죄자를 체포했다고 밝혔다.

2) 실제 감시 체계 활용 사례
2018년 장시성 난창에서 열린 콘서트에서 5만명 관중 가운데서 수배 중이던 남성을 곧바로 찾아내 검거했다. 어떤 사람이 계속 지하철역에 오는데 직원이 아니라면 도둑일 가능성이 있다는 식으로 시스템이 스스로 분석해 이상 신호를 곧바로 공안에 전송한 것이다.
2021년 AI와 안면 인식 기술을 사용해 감정 상태를 드러내는 카메라 시스템이 중국 신장 위구르인들을 대상으로 테스트됐다는 사실이 한 인권 활동가에</td></tr>
</table>

	의해 밝혀졌다. 그는 "중국 정부는 마치 실험실 쥐처럼 위구르인들을 다양한 실험 대상으로 사용 한다."고 말했다. 신장에서 모니터링에 사용되는 빅데이터 시스템은 통합 공동 운영 플랫폼(IJOP)이다. IJOP는 지역의 대량 감시 시스템으로써 전화, 차량 및 신분증을 추적하고 전기 및 주유소 사용을 감시하는 중추 시스템 같은 역할을 하고있다. 심지어 너무 많은 전기를 사용하는 많은 정상적이고 합법적인 활동을 의심스러운 행동의 지표로 취급한다. 어떤 사람들은 추가 심문을 위해 추출되고 구금되거나 투옥된다. 3) 인권 문제 대부분의 사람들은 사생활 보호가 약한 나라에 살고 있다. 정부가 시민과 외국인 모두에 대해 광범위한 감시를 실시하는 미국조차도 개인 정보 보호 법률이 부족하다. 그러나 미국은 개인 정보 보호에 대한 인식이 증가함에 따라 다양한 주와 도시가 캘리포니아 소비자 개인 정보 보호법과 같은 법을 제정하게 된 민주주의 국가이다. 인권 보호의 불안정은 짐바브웨에서 중국에 이르기까지 매우 강압적인 정부가 인권에 대한 공격을 심화시키기 위해 새로운 기술을 적극적으로 모색하는 등 공산주의의 힘이 강한 국가에서 주로 나타난다. 중국에는 효과적인 사생활 보호가 없을 뿐만 아니라 시민 사회, 언론의 자유, 선거도 없다. 공산당은 법 위에 있고, 인터넷을 옥죄고 있으며, 시진핑 체제하에서 반대 의견을 점점 더 용납하지 않고 있다. 이러한 개인정보 침해의 극단적인 끝에는 1,300만 명의 위구르인과 다른 투르크계 무슬림 소수민족이 있는 중국 북서부의 지역인 신장이 있다. 중국 정부는 '폭력 극단주의에 반대하는 강경 시위'에 따라 사실상 그들의 모든 움직임을 추적함으로써 신장의 무슬림 소수자들에 대한 억압을 강화하기 위해 기술을 사용했으며, 이들은 대량의 임의 구금, 강제적인 정치적 세뇌, 이동 제한 및 종교적 억압을 받았다. **3. 시사점** 감시에 대한 정부의 억압은 새로운 유형이 아니지만, 중국 정부의 행태는 우리가 지금 행동하지 않으면 많은 인류의 미래가 될 수 있는 새로운 사회 통제 모델을 제시하고 있다. 과거의 권위주의 국가들과는 달리, 오늘날 억압적인 정부들은 위협적인 디지털 감시 시스템을 마음대로 사용할 수 있다. 그렇기에 정부와 민간 기업 모두 생체 인식 데이터의 수집, 사용 및 저장을 윤리적이고 유용한 도구로써 사용할 수 있도록 제한하는 강력한 규제가 시급하다. 새로운 기술은 종종 사회가 비용과 이익을 이해하고 고민할 시간을 갖기 전에 성급하게 사용된다. 1960년대 후반, 런던 정부는 감시 카메라가 효과적인 범죄 퇴치 도구라며 영구적으로 설치하기 시작했다. 그로부터 약 50년 후 신장에서 인간의 자유를 파괴하고 있는 극적으로 더 강력하고 상호 연결된 시스템으로 이어졌다. 우리는 이제 그것이 국민 감시 강화에 도움이 되었는지를 알 수 있게 되었다. 우리는 너무 늦기 전에 그들의 확산을 막아야 한다.

사고 피해 내역(예상)	1. **피해 대상**: 신장 위구르 자치구의 위구르족 및 기타 무슬림, 중국 국민 2. **피해 내역**: 중국 정부가 카메라와 AI를 이용한 첨단 기술의 감시 네트워크를 통해 숱한 사람들을 억류하고 탄압

SW신기술 안전사고 보고서

No. AI-17

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">아마존 AI 채용 도구의 성별 편향</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hankookilbo.com/News/Read/A2021101409500001667</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 10월 14일</td></tr>
<tr><td>작성자</td><td colspan="3">전창배</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>AI 편향 문제</td></tr>
<tr><td>사고 일시</td><td>2018년 10월 11일</td><td>사고 지속시간</td><td>3년</td></tr>
<tr><td>사고 내용</td><td colspan="4">

1. 관련 AI 기술

인공지능 채용 프로그램은 수만 개의 빅데이터, 자료, 패턴을 학습한 인공지능을 활용하여 응시자의 얼굴, 감정 표정, 색상 변화 등을 분석하여 평가를 진행한다. 또한 음성인식 기반 기술을 적용하여 각각의 사람마다 다른 음향, 음색, 사용 어휘 등을 검토한다.

인공지능를 활용하면 객관성과 공정성을 확보할 수 있다. 인공지능이 평가를 진행하기 때문에 응시자 모두에게 동일한 시간이 주어지며, 인간의 선입견이 배제되기 때문이다. 또한 채용 과정 시 발생하는 시간과 비용을 절감할 수 있다는 장점이 있다.

2. 사고 경위

1) 보도 내용

2018년 10월 11일, 글로벌 기업 아마존에서 인공지능 채용 프로그램 논란이 발생했다. 아마존에서 자체 인공지능 채용 프로그램을 개발했는데, 실제 채용에 적용하기 전 최종 시뮬레이션 과정에서 남성 지원자가 여성 지원자보다 평균적으로 더욱 높은 점수를 받는 편향 현상이 발견되었다. 원인은 아마존의 직원 구성으로 밝혀졌다. 아마존은 기본적으로 IT 기업으로써 개발직군이 70%를 차지했으며 개발직군의 남자 직원이 여성 직원보다 압도적으로 많은 인원이였다. 그에 따라 남성 직원 중에 고성과자가 많았다. 인공지능 채용 프로그램은 아마존 내부 직원들의 데이터를 학습하고 기준으로 삼았기 때문에 해당 모수를 곧 채용의 기준이 되었다. 2021년 2월 11일, 결국 아마존은 데이터를 수정하거나 알고리즘을 수정하는 것으로는 한계가 있다고 판단하여 인공지능 채용 프로그램을 폐기하기로 결정했다.

2) 아마존의 대응

아마존의 채용 담당자들은 신규 채용을 검색할 때 툴이 생성하는 추천 사항을 살펴보았지만, 그 순위에만 의존한 적은 없었다고 주장했다. 아마존은 이 기술에 대해 언급을 피하면서도 “관계자들이 후보를 평가하기 위해 이 도구를 사용한 적이 없다”라고 말했다. 아마존은 그 이상 자세히 설명하지 않았

</td></tr>
</table>

	다. 채용 담당자들이 채용 엔진에서 생성된 추천서를 살펴본 것에 대해선 이의를 제기하지 않았다. 현재, 여성들은 아마존 노동력의 40%를 차지하고 있다. 3) 채용 알고리즘과 젠더 편향 채용 알고리즘 자체가 아마존뿐만 아니라 많은 기업에 걸쳐 사용되고 있지만 여전히 젠더 편향, 즉 여성이 지배하는 인사부의 역할과 기술의 평가절하를 야기하고 있다. 미국 노동부(Department of Labor)에 따르면 여성이 인사 관리 직무의 거의 4분의 3을 차지한다. 이것은 직장에서 전반적인 여성 대표자들에게 좋은 이야기이다. 그러나 그 차이에는 또 다른 종류의 성 편견이 존재한다. 인사 관리 직무는 여성적인 역할이라는 인식이 있다. 인사 관리는 여성의 직업이라는 인식이 지속되고 있다. '사람을 기반으로 하고, 부드럽고, 공감적이며, 직원들이 문제를 해결하도록 돕는' 이러한 이미지는 대부분 여성들에 의해 전형적인 양육자로 보이게 한다. 오늘날에도 이러한 여성들이 "더 부드럽다"라는 인식은 직원 능력에서 벗어나 남성은 불리하다고 판단되고 있다. 채용에서 AI 통합을 추구했던 아마존과 다른 회사들은 채용 과정을 간소화하는 방향으로 나아가려 했다. 그러나 사람 기반 프로세스를 자동화하는 것은 직관이나 관계와 같이 기계적으로 재현하기 어려운 사람 기반 기술을 무시하는 것을 보여준다. AI가 채용을 진행하는 것은 마치 쇼핑객들이 상품을 평가하는 것과 같다는 의견이 있다. **3. 시사점** 사회 구조와 사회 문화가 존재함에 따라 데이터의 편향이 발생할 수밖에 없다면 이를 학습하는 인공지능 역시 편향될 수밖에 없다. 따라서 현재 기술 수준에서 완벽한 인공지능의 성능을 바라는 것보다 그 한계점을 먼저 인정하는 것이 기술 활용에 도움이 될 것이다. 우리는 AI의 편향성을 방지하기 위해 기획부터 출시, 운영과정까지 모든 과정에서 지속적으로 검증하고, 보완하고, 또 검증하고, 보완해야 한다. 결과에 대해 충분한 시뮬레이션을 반복하고 그렇게 검증 과정에서 나온 오류들을 수정하고, 또다시 시뮬레이션하고 그 결과를 다시 검증하고 보완하는 절차를 반복해야 한다. 정부와 기업 및 기관들 모두 협력하여 이러한 AI 편향성을 테스트하고 검증할 수 있는 시스템을 개발하고 보급하는 것이 필요하다. 또한 그러한 편향성을 검증하고 가이드할 수 있는 중립적인 기구를 설치하여 우리 기업들을 지원하는 것도 좋은 방안이 될 수있다.
사고 피해 내역(예상)	**1. 피해 대상:** 채용 과정 중 불이익을 받은 지원자들 **2. 피해 내역:** 아마존 인공지능 채용 프로그램에서 남성 지원자가 여성 지원자보다 더욱 높은 점수를 받음으로써 여성 지원자들에게 불이익이 발생

SW신기술 안전사고 보고서

No. AI-18

<table>
<tr><td>사고 분야</td><td colspan="4">AI</td></tr>
<tr><td>사고명</td><td colspan="4">범죄예측 프로그램 COMPAS의 인종 편향</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hani.co.kr/arti/science/future/914757.html</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2019년 10월 27일</td></tr>
<tr><td>작성자</td><td colspan="3">곽노필</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>AI 편향 문제</td></tr>
<tr><td>사고 일시</td><td>2018년 1월 17일</td><td>사고 지속시간</td><td>진행 중</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 AI 기술
세계적으로 사법 시스템에 인공지능의 활용을 권고하고 적극 도입하는 나라가 증가하고 있다. 주로 전략적 재판 보조 도구로써 활용되고 있는데, 유사한 분쟁 결과 데이터들을 이용해 통계적 분석을 진행하고 판단하는 역할을 해준다. 민사 분쟁 결과, 민사절차법상 대체적 분쟁 해결 절차, 온라인 분쟁 해결 절차 등 더욱 다양한 방면에 도입할 것이라고 예상되고 있다.
AI 재판 지원시스템 컴파스(COMPAS)는 미국 스타트업 ‘노스포인트사’에서 만든 피고인의 데이터를 종합하여 재범 가능성을 예측하는 인공지능이다. 법원에서 형사 재판 피고인에 대해 형을 선고할 때 AI의 알고리즘 자료를 근거로 삼는 식으로 활용하는 것이다. 컴파스는 재판의 효율성과 일관성을 증가시켜 주기 때문에 미국 법원에서 과거에 암묵적으로 사용했으며 실제 합법화한 판결을 내린 사례도 있다.

2. 사고 경위
1) 보도 내용
COMPAS를 분석한 결과, 흑인 피고인이 백인 피고인보다 재범 위험성이 높은 것으로 잘못 판단될 가능성이 훨씬 높은 반면, 백인 피고인은 흑인 피고인보다 낮을 가능성이 더 높은 것으로 판단하고 있다는 사실을 발견했다. 플로리다 브로워드 카운티에 있는 1만 명 이상의 형사 피고인들을 조사했고 그들의 예상 재범률을 2년 동안 실제로 발생한 비율과 비교한 결과, COMPAS 도구에 의해 예측된 재범 위험 범주와 점수가 매겨진 후 2년 동안 피고인의 실제 재범률을 비교한 결과 점수는 가해자의 재범률을 61%로 정확하게 예측했지만 폭력적 재범률을 20%로 예측한 경우에만 정확한 것으로 나타났다. 누가 재범할지 예측할 때, 알고리즘은 흑인과 백인 피고인의 재범률을 거의 같은 비율로 정확하게 예측했지만(백인 피고인 59%, 흑인 피고인 63%) 매우 다른 방식으로 실수를 했다. 분석 결과, 과거 범죄, 재범, 연령, 성별 등을 통제할 때에도 흑인 피고인이 백인 피고인보다 위험 점수가 더 높게 배정될 가능성이 45% 더 높았기 때문이다.

2) 세 가지의 문제</td></tr>
</table>

재범 위험 평가 과정은 주체의 자유에 대한 적절한 제한을 결정하는 데 도움을 주기 위한 것이다. 인공지능 분야를 중심으로 기술 혁신이 이뤄지면서 이제 AI 기술을 기반으로 한 재범 위험 평가 도구가 잘 개발돼 형사사법제도에 활용되고 있다. 형사 사법 시스템에서 재범 위험 평가에 AI 기술의 사용이 증가하는 것은 많은 법적 문제를 제기한다.
첫 번째는 알고리즘이 인종차별을 하는 것이다. 이 문제는 AI 위험 평가 도구를 훈련할 때 과거 데이터를 사용하는 것에서 발생한다. 위험 평가 중 역사적 편향을 영구화하는 경향이 있기 때문이다.
두 번째 문제는 판결 결정에서 판사의 AI 위험 평가 사용의 합법성과 관련이 있다. 일반 인구의 데이터를 기반으로 한 AI 위험 평가는 정확한 정보를 기반으로 한 개인화된 문장에 대한 권리를 죄수로부터 박탈하는 경향이 있기 때문이다.
셋 번째로 AI 위험 평가 도구에 사용되는 방법론의 독점적 특성에 문제가 있다. 특히 형사처벌에 이의를 제기하는 범죄자들이 이러한 독점적 영업비밀에 대한 접근을 추구하는 경우가 이에 해당한다.

3. 시사점

데이터 분석에 AI 기반 도구를 사용함으로써 발생하는 체계적이고 인종 기반 편향은 인종으로 구분된 사회의 특정 계층의 개인에게 불공평한 결과를 초래한다. 인공지능은 빅 데이터에 의존한다. 인공지능 알고리즘이 예측을 할 수 있도록 훈련시키려면 방대한 양의 데이터가 필요하다. 우리는 이 도구를 활용하기 전 항상 데이터가 충분하게 쌓였는지 재차 검증하고 의심해볼 필요가 있다.
기술 발전은 이론적으로 환영할 만한 발전이다. 그러나 실제로는 자동화의 측면이 사법 제도를 형사 피고인들에게 덜 공평하게 만들고 있다. 문제의 근본은 자동화된 형사사법 기술이 대부분 개인 소유로 되어 있고 이익을 위해 판매되고 있다는 것이다. 개발자들은 그들의 기술을 영업 비밀로 삼는 경향이 있다. 그들은 형사 소송이나 가석방 심리가 통제된 상황에서 보호 명령 하에서도 형사 피고인과 그들의 변호사에게 그들의 알고리즘이 어떻게 작동하는지에 대한 세부 사항을 공개하는 것을 종종 거부한다. 이건 형사사법제도의 새롭고 골치 아픈 특징이고 개발자에게도 피해자 모두에게 좋은 해결 방법이 필요할 것이다.
인공지능 기술은 계속해서 우리의 사법 시스템에 혁명을 일으킬 것이다. 만약 적절하게 사용된다면, 우리 형사사법제도의 효율적이고 효과적인 행정을 크게 향상시킬 수 있을 것이다. 그러나 형사 사법 시스템에서 AI를 사용하는 것은 해결해야 할 몇 가지 심각하고 새로운 법적 문제를 제기한다. 이러한 기술이 피고인과 범죄자의 권리에 미치는 부정적이고 차별적인 영향을 완화할 수 있는 프레임워크를 개발하는 궁극적인 목표를 가지고 이러한 법적 문제를 연구하는 것이 중요하다.

사고 피해 내역(예상)	1. **피해 대상**: AI 재판 지원시스템 컴파스(COMPAS)가 올바르게 판단하지 못한 흑인 피해자들 2. **피해 내역**: 흑인은 재범 위험으로 잘못 식별되는 경우가 많고 백인은 위험이 아닌 것으로 잘못 식별되는 경우가 많은 것이 발견

SW신기술 안전사고 보고서

No. Metaverse-01

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">미성년자 대상 디지털 성범죄</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.donga.com/news/Society/article/all/20220414/112869486/1</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 4월 14일</td></tr>
<tr><td>작성자</td><td colspan="3">김윤이, 이채완</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>한국(제페토)</td><td>사고 유형</td><td>성범죄</td></tr>
<tr><td>사고 일시</td><td>2021년 4월 ~ 2022년 3월</td><td>사고 지속시간</td><td>1년</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 메타버스
제페토(ZEPETO)는 네이버 자회사인 네이버제트의 대표적인 국내 메타버스 서비스로 2018년 출시 이후 전세계 2억명이 사용하고 있는 인공지능을 접목한 3D 아바타를 기반으로 한 가상 세계 플랫폼이다.
제페토는 가상 세계에서 아바타를 통한 라이프로깅 서비스를 제공한다. 이용자들은 아바타를 통해 가상공간을 돌아다니며 친구 아바타와 사진을 찍고, 아바타를 주인공으로 한 미니 드라마를 만들기도 한다.

2. 사용자 특성
국내에서 제페토 이용자는 7~12살이 50.4%, 13~18살이 20.6%를 차지하고 있는 등 미성년자의 이용률이 높은 편이다.

3. 사고 경위
1) 위장: 가해자는 제페토에 가입해 디지털 캐릭터인 자신의 아바타를 미소년 같은 외모로 치장한 뒤 자신을 20대 초반의 대학생으로 소개하며 여성 미성년자들에게 접근하였다.
2) 관계 형성: 가해자는 대화를 시도하면서 피해자들과 길게는 1~2개월간 연락하면서 친분을 쌓았고, 피해자들의 환심을 사기 위해 메타버스 내 유료아이템과 기프티콘 등의 선물을 지속적으로 보내며 관계를 돈독하게 구축하고 신뢰를 형성했다.
3) 범죄행위: 피해자들과 지속적인 친분을 쌓은 뒤 1년간 자신의 아바타에 관심을 보인 아동, 청소년을 대상으로 성착취 목적 대화를 하고 성착취물을 제작한 혐의를 받고 있다. (친분을 쌓은 뒤 성범죄 행위 진행)
4) 가해자의 행동 특성 : 가해자는 제페토에서 총 5개의 계정을 사용하면서 계정이 정지된 뒤에도 새로운 계정으로 재가입해 피해자인 아동, 청소년들에게 접근하였다.</td></tr>
</table>

	5) 검거 방법 : 경찰은 제페토 내에서 미성년자를 대상으로 성범죄가 일어난다는 첩보를 입수 한 뒤 수사에 착수했으나, 제페토는 사살 상 익명으로 가입할 수 있기 때문에 바로 검거가 어려워 가해자의 혐의를 특정한 경찰이 위장 신분으로 접근하는 위장수사 기법을 활용해 검거에 성공하였다. 위장수사를 통해 범행 수법까지 파악했으며, 추적 한 지 2주만에 가해자를 주거지에서 검거 한 뒤 구속돼 검찰로 넘겨졌다. **4. 사고조사 결과 및 조치** 검거 후 확인 결과 가해자는 20대 대학생이 아닌 30대 중반의 일용직 남성으로 확인되었으며, 아이들이 주로 사용하는 언어나 대화방식, 트렌드, 유행에 능통했던 것으로 파악됐다. 가해자는 피해자들에게 몸을 찍은 사진과 영상을 보내달라고 요구하였으며 받은 사진을 이용해 성 착취물을 제작하였다. 또한 메신저를 통해 전화를 걸어 성적 대화를 하였으며 자신의 신체를 찍은 영상을 피해자들에게 직접 보내는 등의 행위를 한 것으로 확인되었다. 가해자가 제작한 성착취물은 외부로 유출된 정황이 포착되지는 않았으며, 경찰은 2차 피해를 막기 위해 모든 성착취물을 삭제하고 제페토에 가해자의 계정 폐쇄를 요청하였다. 아울러 피해 아동·청소년들을 상대로 보호기관 연계 등 보호 활동도 전개 중이며 향후 메타버스 공간상 디지털 성범죄에 대한 단속을 강화할 것이라고 밝혔다. **5. 사고에 대한 제페토의 입장** 제페토는 AI를 이용해 금칙어(성적 수치심을 일으킬 수 있는 언어나 욕설)를 걸러내고 메타버스 안에서의 윤리 의식을 고취하는 캠페인을 벌이고 있다. 구체적으로는 '상대방의 외모 평가는 자제해 주세요'와 같은 메시지를 보내는 한편 음란성 발언, 불건전 행위 등 가이드라인을 지키지 않으면 채팅을 제재하는 방식이다. 사용자의 경험과 메타버스 상 행동 양식을 주의 깊게 지켜보며 플랫폼에 적합한 해결 방식을 찾으려는 노력도 계속되고 있다. 제페토 운영사 네이버제트는 보호 시스템 모니터링을 통해 계정 정지 등 자동 제재 시스템을 가동하고, 신고 담당 모니터링 요원을 추가 채용하고 있다. 한국청소년상담복지원과의 협력 캠페인을 진행하고 부모·보호자를 위한 안내 가이드 신설하는 등 이용자 보호에 힘쓰고 있다는 입장이다.
사고 피해 내역	**1. 피해 대상:** 초등학교 저학년 어린이부터 고등학생까지 여성 아동 및 청소년 11명 **2. 범행 수법:** 선물을 이용해 피해자들의 환심을 산 뒤 자신의 신체 사진을 먼저 찍어 보낸 후 피해자들의 신체 사진을 요구한 온라인 그루밍(심리적 지배)을 통해 성 착취물을 제작하여 보관하고 메신저를 통해 성적인 대화를 지속 **3. 피해에 대한 조치:** 정신적 피해를 입은 아동, 청소년을 상대로 보호기관 연계

SW신기술 안전사고 보고서

No. Metaverse-02

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">현실과 비슷해지는 디지털 성범죄</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">1) https://n.news.naver.com/article/050/0000059957
2) https://www.seoul.co.kr/news/newsView.php?id=20220528500021</td></tr>
<tr><td>보도 날짜</td><td colspan="3">1) 2022년 2월 21일
2) 2022년 5월 28일</td></tr>
<tr><td>작성자</td><td colspan="3">1) 정재희
2) 김재현</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국(호라이즌월드)</td><td>사고 유형</td><td>성범죄</td></tr>
<tr><td>사고 일시</td><td>2021년 12월 ~</td><td>사고 지속시간</td><td>계속</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 메타버스
호라이즌 월드(Horizon Worlds)는 메타(Meta, 구 페이스북)에서 2021년 12월에 출시한 소셜 미디어 플랫폼으로 VR을 기반으로 가상현실 소셜 공간을 제공한다.
VR 헤드셋인 오큘러스를 착용하고 ‘호라이즌’이라는 가상현실에서 한번에 최대 20개 아바타가 함께 모여서 탐험을 즐기고 이벤트에 참여하거나 대화 등의 상호작용을 비롯해 각종 게임도 즐길 수 있다.

2. 메타버스 특성
VR기기가 있는 이용자는 애플리케이션(응용프로그램)을 실행해 다리 없이 상반신만 있는 아바타를 만들고 다른 이용자와 교류하거나 게임을 할 수 있다. 오큘러스 VR 장비를 착용해야지 사용 가능한 것이 다른 메타버스 플랫폼과의 차이점이다.

3. 사고 경위
1) 2021년 12월
① 호라이즌 월드는 베타 테스트 과정에서부터 성폭행, 성추행 피해 문제가 불거졌다. 아동을 위한 메타버스 기술 연구 업체인 카부니의 니나 파텔 부사장이 2021년 12월 22일에 본인이 호라이즌 월드에서 경험한 성추행 피해를 온라인에 공개한 내용에 따르면, 당시 파텔 부사장의 아바타는 가상현실에서 접속한 지 60초도 채 되지 않아 3~4명의 남성 아바타에 둘러싸여 집단 성폭행과 성희롱을 당했으며 너무 빠르게 일어난 사건에 속수무책으로 당할 수밖에 없었다고 밝혔다.
② 위의 사건이 벌어진 뒤 호라이즌은 내부 검토를 통해 아바타 사이의 거</td></tr>
</table>

	리가 약 4피트(120cm)로 유지되도록 하는 개인 경계선(Personal Boundary) 즉, 아바타 간 거리두기 기능을 도입해 한 아바타가 다른 아바타에 접근하더라도 개인 경계선 내로 진입할 수 없는 기능을 추가했다. 2) 2022년 5월 ① 2021년 12월에 이어 비영리단체 '섬 오브어스(Sum of Us)'는 여성연구원의 체험을 바탕으로 '메타버스: 중독성 있는 콘텐트의 또 다른 시궁창'보고서에 호라이즌을 테스트하면서 어떤 일을 경험 했는가에 대한 내용을 담았다. 이 보고서에 따르면 여성 아바타로 호라이즌 월드에 접속 후 한 시간 만에 파티가 열린 방으로 안내된 뒤 다른 이용자로부터 성폭행을 당했으며 또 다른 이용자는 보드카 술병을 들고 이 장면을 촬영하는 흉내를 내기까지 했다. 다른 이용자가 아바타를 만지면 장착하고 있는 VR기기가 진동하면서 물리적으로 불편한 느낌이 실제 사용자에게 전달되었다. 1)의 사건 이후 호라이즌 월드는 아바타 간 거리두기 기능을 도입했지만 사용자가 이를 임의로 해지할 수 있다. 보고서의 연구원 역시 기본 경계 기능은 기본적으로 활성화되어있지만 다른 사용자의 권유로 이 설정을 해제했다고 밝혔다. **4. 사고조사 결과 및 조치** 섬오브어스는 메타가 사용자들의 문제 행동을 통제할 관리자를 두고 있지만, 그 숫자가 부족하고 서비스에 대해 안내를 할 뿐 실질적으로 개입하지 않는다고 지적했다. 또 사용자의 접근을 막거나 음소거 하는 등의 기능이 있지만, 문제를 일으킨 사용자를 처벌하지 않는 등 한계가 있다고 주장했다. **5. 사고에 대한 호라이즌의 입장** 2021년 12월 사고 공개 호라이즌은 개인 경계 기능, 즉 아바타 간 거리두기 기능을 추가하여 친구가 아닌 사람이 자신의 아바타에서 약 4피트 이내고 들어오지 못하도록 하는 기능을 추가했다. 2022년 5월 사고 후에는 원치 않는 접촉을 피할 수 있도록 안전 기능이 기본으로 설정이 됐다며 모르는 이용자에 대해서는 이 기능을 해제하지 않는 것이 좋다고 설명하며 계속 연구하고 조치를 취할 것 이라고 말했다.
사고 피해 내역	**1. 피해 대상:** 주로 여성으로 보이는 아바타를 향한 성희롱과 성폭행 **2. 피해 내역**: VR은 본질적으로 마음과 몸이 가상 경험을 실제와 구별할 수 없도록 설계되어 생리적이고 심리적인 반응은 마치 현실에서 일어난 일처럼 동일하게 느껴져 현실과 유사한 정신적, 신체적 피해가 우려되며 메타버스 서비스가 현실과 가까워질수록 피해자 입장에서 피해의 현실성이 함께 올라감 **3. 피해에 대한 조치:** 미국의 소비자 단체도 메타버스 내 성폭력 등 각종 범죄가 제지받지 않고 빈번히 발생한다며 정부와 플랫폼이 적극적인 규제에 나서야 한다고 촉구하고 있으며 현행법 상 캐릭터를 통해 성희롱이나 성행위를 시키더라도 이를 제재할 규정은 없어 이제 대한 문제 제기가 일어나고 있다.

SW신기술 안전사고 보고서

No. Metaverse-03

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">메타버스 공간 내 지적재산권 침해</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.econovill.com/news/articleView.html?idxno=536342</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021.06.11</td></tr>
<tr><td>작성자</td><td colspan="3">민단비</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국(로블록스)</td><td>사고 유형</td><td>저작권 침해</td></tr>
<tr><td>사고 일시</td><td>2021년 6월 9일</td><td>사고 지속시간</td><td>3개월</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 메타버스
로블록스(Roblox)는 전세계 1억 5천만명이 사용하고 있는 메타버스 플랫폼으로 게임 속의 아바타를 통해 일상생활을 수행하며, 사용자가 게임 콘텐츠를 쉽게 개발이 가능하다.

2. 메타버스 특성
이용자가 직접 게임을 제작할 수 있는 ‘로블록스 스튜디오(Roblox Studio)’라는 창작 도구를 제공하고 있으며 이 스튜디오를 이용하면 프로그래밍 지식이 부족해도 간단하게 게임을 제작하고, 로블록스 플랫폼에 등록 및 판매로 수익 창출이 가능하다. 로블록스는 미국의 16세 미만 청소년 중 55% 이상이 가입해 인기 플랫폼으로 자리 잡고 있다.

3. 사고 경위
1) 전미음악출판협회(National Music Publishers Association, 이하 NMPA)는 2021년 6월 9일 로블록스가 자사 플랫폼에서 음악 저작권자들에게 비용을 지불하지 않고 음악을 불법으로 사용하고 있다는 이유로 로블록스를 상대로 약 2억 달러(한화 2,230억원) 규모의 손해배상 청구소송을 캘리포니아 지방법원에 제기했다.

2) NMPA 측의 소송 사유
NMPA는 로블록스가 저작권에 대한 이해가 부족한 어린 이용자들을 이용하여 로블록스 플랫폼에 음악을 업로드 할 때마다 비용을 지불하도록 하여 수억달러를 벌어들였지만, 이용자들의 반복적인 저작권 침해를 금지하거나 이용자들의 행위에 대한 위험성을 경고하는 어떠한 조치도 하지 않았다고 주장하였다. 사용자들은 로블록스에서 제공하는 음악 라이브러리에서 음악을 선택해 비디오 게임이나 컨텐츠 제작에 활용할 수 있다. 이 과정에서 사용자는 일단 자신</td></tr>
</table>

	의 컴퓨터에 음악을 다운로드 한 후, 게임이나 콘텐츠용으로 다시 사이트에 업로드하게 되는데, 이때 '로벅스'라는 로블록스 내 가상화폐의 추정가가 나오며 음원파일의 크기가 클수록 로벅스를 통해 업로드 하는 비용도 커지게 된다. 협회 측은 로블록스가 사용자들이 음악을 업로드 할 때 마다 수수료 방식으로 수익을 창출하면서, 저작권 라이선스 계약을 체결하지 않고 음반 제작사와 창작자들에게 저작권료를 지급하지 않고 있다고 하였다. 3) 소송의 쟁점 쟁점은 로블록스가 미국 저작권법의 온라인서비스 제공자의 면책조항에 따라 현재 커뮤니티 규정과 필터링 시스템만으로도 사용자 침해가 발생하는 경우 면책이 될 수 있는가이다. 이에 대해 NMPA 측은 로블록스가 미국 저작권법의 온라인서비스 제공자의 면책조항을 의도적으로 오독해 책임을 회피하고 있다고 주장하며 지속적인 저작권 침해, 그리고 그 침해를 기반으로 하여 부당한 이익을 취하고 있음에도 불구하고, 단순히 커뮤니티 규정만을 공표하는 정도로는 면책 대상이 될만한 충분한 조치가 아니라고 하였다. **4. 소송에 대한 로블록스의 입장** 로블록스는 로블록스 플랫폼의 작동 방식에 대한 이해 부족에서 비롯된 소송이라고 설명했다. 창작자 커뮤니티가 스스로 만들어가는 플랫폼으로서 인디뮤지션, 작곡자, 음반사 및 음반 제작사들의 지식재산권 보호에 적극적이며, 로블록스 커뮤니티 회원들은 커뮤니티 규정을 준수하고 있다고 반박했다. 더불어 미국 저작권법(DMCA)에 따라 저작권 침해물들이 발견될 경우 즉시 삭제하고, 상습 침해자들은 플램폼의 사용을 엄격하게 제한하고 있다면서 오히려 음반 제작사 등과 지속적으로 제휴관계를 맺으며 뮤지션을 포함한 예술 창작자들에게 새롭고 창의적인 수익 창출 기회를 가상세계에서 만들어 나가고 있다고 주장했다. **5. 진행 결과** 이 소송전은 비슷한 서비스를 제공하는 메타버스 플랫폼과 저작권자들의 권리관계와 계약 문제를 정리해줄 것이라는 기대 속에 판결 결과가 주목받고 있었으나, 2021년 9월 27일 두 당사자가 분쟁 사항에 대한 합의를 결정하여 법원의 판단을 확인할 수 없게 되었다. NMPA와 로블록스는 NMPA 소속의 모든 음반사들이 각각 자유롭게 로블록스와 라이선스 계약을 체결하고, 자유롭게 계약 조건에 대해서 합의할 수 있는 옵트인(Opt-In) 방식에 대한 합의를 하는 것으로 마무리 되어 로블록스는 향후 NMPA 회원사들과 개별적인 라이센싱 계약을 추진해 나가게 될 전망이다.
사고 피해 내역	**1. 피해 대상**: 미국 주요 음반사 **2. 피해 내역**: 로블록스가 플랫폼 안에서 라이선스 계약 없이 가상 음악 재생 장치를 통해 음악저작물을 무단으로 이용했다는 이유로 전미음악출판협회가 로블록스를 상대로 약 2억달러의 소송 제기

	3. 결과: 당사자간의 합의로 마무리 되었으며 로블록스는 향후 NMPA 회원사들과 개별적인 라이센싱 계약을 추진해 나가게 될 전망

SW신기술 안전사고 보고서

No. Metaverse-04

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">가상부동산 서비스 중단으로 인한 유료 구매자의 금전적 손실</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://n.news.naver.com/mnews/article/277/0005063787?sid=101</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 3월 25일</td></tr>
<tr><td>작성자</td><td colspan="3">최동현</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>한국(세컨서울)</td><td>사고 유형</td><td>서비스 중단</td></tr>
<tr><td>사고 일시</td><td>2021년 12월</td><td>사고 지속시간</td><td>3일</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 메타버스
세컨서울(2nd Seoul)은 실제 서울 지도를 타일 수백만개로 나누고, 이를 이용자가 보유할 수 있도록 설계한 메타버스 가상부동산 거래 플랫폼이다. 현실 세계의 부동산 소유권과 관계가 없지만 메타버스 속에서는 대체불가토큰(NFT)을 통해 특정 지역의 소유권을 보장한다. 엔비티의 100% 자회사인 엔씨티마케팅이 자체 개발하였으며 유사한 메타버스 부동산 플랫폼으로는 어스(Earth)2가 있다.

2. 메타버스 특성
메타버스 가상부동산 서비스 유형은 크게 두 가지다. ‘더 샌드박스’나 ‘디센트럴랜드’처럼 가공의 세계를 만들거나,‘어스2’혹은 ‘세컨서울’처럼 실제 국가나 도시를 가상공간에서 그대로 구현하는 방식이다. 가상부동산 매매는 땅을 타일 단위로 나누거나 청약을 통해 건물을 분양하는 등의 방식으로 이뤄지고 있다. 세컨서울 서비스는 가상부동산 주인인 '타일오너'와 특정 타일에 광고를 하려는 '소상공인', 세컨 서울에 실제 위치 정보와 소비정보를 제공하는 '세컨시민'이 활동하면서 가치를 만들어 낸다고 회사는 설명했다.

3. 사고 경위
1) 사전 오픈 및 런칭
세컨서울은 2021년 12월 29일 론칭 하였다. 론칭 전 사전 신청 인원수가 한달도 채 안돼 25만명에 이를 정도로 큰 기대를 모았다. 사전 신청은 일종의 청약과 유사한 개념으로 서비스 오픈 시 가상 세계에서 서울 부동산을 받을 수 있는 자격을 얻을 수 있다. 현실에서 지가가 높은 지역인 광화문, 강남역 일대나 용산구 한남더힐, 서초구 반포자이 등 고가 주거지역이 일찌감치 신청이 마감됐다. 론칭 당시 타일 1개당 액면가는 1만원 이었고, 서울 전역을 가상부동산 타일 694만개로 나누었다. 소유권 인증은 NFT 발급을 통해 이</td></tr>
</table>

	루어지고, 향후 가상부동산을 통해 투자, 임대, 광고 수익을 얻도록 하였다. 2) 서비스 중단 이러한 흥행 성공에도 불구하고 2021년 12월 31일 세컨서울은 출시 이틀 만에 서비스 점검에 돌입했고 갑자기 서비스가 중단됐다. 세컨서울 측은“결제 및 보안 프로세스 안정성 확보 측면에서 일부 이슈 사항들을 발견함에 따라 지속적인 서비스 유지에 무리가 있을 것으로 판단했으며, 세컨서울의 서비스 제공 주체를 모회사인 주식회사 ‘엔비티’로 변경 후 본격적인 정식 서비스 준비에 역량을 집중할 것”이라고 입장을 밝히며, 유료로 타일을 구매한 고객에 대해서는 수수료 500원을 포함한 1만500원 전액을 환불해 주기로 하였다. 유료 구매한 유저가 상당수 있었을텐데 베타서비스였다며 일방적으로 서비스를 중단한 뒤 양해해 달라고 한 입장문이 아쉽다는 반응과 함께 현실과 다소 괴리감이 있는 게임 스타일의 메타버스가 아닌 현실 생활과 밀접하게 영향을 주고받는 서비스라 성장성이 기대된다는 반응도 있었다. 3) 소송 이후 엔비티는 엔씨티마케팅이 모회사인 엔비티와 상의 없이 세컨서울 거래 서비스를 불완전한 상태로 무단 출시했다는 이유로‘세컨서울’ 사업을 준비해 온 자회사 엔씨티마케팅의 전 대표이사와 다른 이사진 3인, 외부용역 개발자 4인에 대한 고소장을 접수했다. 엔비티 측은 엔씨티마케팅이 세컨서울 서비스를 개발 단계에서 불완전 상태로 무단 론칭했으며, 이후 이틀 만에 서비스를 종료하는 과정에서 ▲전자금융거래법 위반 ▲자본시장법 위반 ▲업무상 배임 등이 드러났다고 주장했다. **4. 소송에 대한 앤씨티마케팅 전 대표 입장** 소송 및 보도에 대해 엔씨티마케팅 전 대표 측은 “엔씨티마케팅 대표이사 및 임직원의 업무상 배임 등 비위 혐의는 2022년 4월 현재 수사 중인 사안으로서 사실로 밝혀진 바 없으며, 전 대표가 주식을 처분한 2022년 1월 25일은 엔비티가 ‘세컨서울’ 서비스를 중단하고 전 대표를 해임한 2021년 12월 29일로부터 약 1개월이 지난 시점으로, 그 처분단가도 21,613원일 따름이라고 밝혔다. **5. 진행 상황** 세컨서울 서비스 관계자인 엔씨티마케팅 대표이사 및 임직원의 업무상 배임 등 비위 혐의는 2022년 4월 현재 수사 중인 사안 중이다. 엔비티와 세컨서울 서비스 관련자 양 측의 서비스 사용권 및 소송 문제로 기존 세컨서울 서비스는 폐기하고 메타서울이라는 신규 서비스를 오픈 할 예정이라고 엔비티 측은 밝혔다.
사고 피해 내역	**1. 피해 대상:** 세컨서울 서비스 이용자 중 유료로 타일을 구매한 이용자

	2. **피해 내역**: 세컨서울 서비스 중단 이후 엔비티와 세컨서울 서비스 관계자들의 소송으로 인해 유료로 타일을 구매한 이용자들이 보상을 받지 못하게 됨

SW신기술 안전사고 보고서

No. Metaverse-05

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">가상세계 과몰입으로 겪는 신체 부상</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">http://it.chosun.com/site/data/html_dir/2022/02/04/2022020402098.html</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022.02.06</td></tr>
<tr><td>작성자</td><td colspan="3">이은주</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td></td><td>사고 유형</td><td>신체 부상</td></tr>
<tr><td>사고 일시</td><td></td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 메타버스
메타버스 가상 세계에서 시간을 보내는 이용자가 늘어나면서 가상 세계에 지나치게 과몰입한 이용자가 겪는 부상 사고가 일어나고 있다.
2022년 2월 4일 월스트리트저널(WSJ) 등 외신에 따르면 VR 헤드셋을 이용해 메타버스에 진입하는 이용자가 크게 늘었다. 실제 미국 내 VR 헤드셋 판매는 2021년에 전년도 대비 70% 증가한 790만대를 기록했는데, 이는 헤드셋을 통해 가상공간에 접속해 게임이나 가상운동을 즐기는 이들이 늘어났기 때문이다.
문제는 VR기기를 이용해 가상 세계에 접속하는 이용자가 많아지면서, 현실 세계에서 각종 신체적 부상 사건 사고가 빈번히 벌어지고 있다는 데 있다. VR을 이용한 메타버스 공간은 현실 세계에서 이용자의 시각과 청각 등 기능을 차단시켜 공간감을 잃게 만들 수 있기 때문에 신체 부상의 위험성도 있다.

2. 부상 유형
외신에 따르면 가상공간에서 운동에 몰입하다가 벽에 부딪혀 큰 부상을 입거나, 물건에 걸려 넘어지는 사례가 빈번하게 일어나고 있다. 역동적인 게임에 몰입하다가 균형감을 잃고 쓰러져 응급실로 실려가는 사건이나, 어깨 탈구도 빈번했다. 인터넷 커뮤니티에서는 VR 게임을 하다가 TV를 부수거나 손으로 벽을 치는 사고도 발생하고 있다.

3. 업계의 대책
이에 관련 기업들은 메타버스 이용 시 이용자의 안전을 보장하기 위한 각종 기능을 추가하고 있다. 메타는 VR 서비스에 안전 이용 공간을 설정하고, 이용자가 이를 벗어나면 가상공간 안에 경고 표시가 나타나도록 하는 기능을 추가했다. 또 메타버스에 접속한 이용자 주변으로 사람이나 애완동물 등이 나타날 때 이를 경고하는 기능도 넣었다. 소니는 VR 헤드셋 내에 안전 이용 가이드</td></tr>
</table>

	를 넣어 이용자들이 참조하도록 했다. **4. 전문가 입장** 전문가들은 지나친 메타버스 몰입을 경계하고 나섰다. 제이 킴 오르곤 주 대학(Oregon State University) 부교수는 "메타버스를 이용하는 과정에서 유해한 신체적 부상을 피하려면 가급적 자주 휴식을 취해야 한다"고 권했다.
사고 피해 내역	1. **피해 대상**: 메타버스를 이용하는 일부 사용자 2. **피해 내역**: 메타버스를 이용하는 과정에서 가상현실에 대한 과몰입으로 인해 신체 부상을 겪는 사용자 발생

SW신기술 안전사고 보고서

No. Metaverse-06

<table>
<tr><td>사고 분야</td><td colspan="4">AR/Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">AR 게임 포켓몬 고로 인해 발생한 교통사고 및 안전사고</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://zdnet.co.kr/view/?no=20170217093800 외</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2017년 2월 17일 외</td></tr>
<tr><td>작성자</td><td colspan="3">이도원 기자 외</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국, 일본, 한국 등</td><td>사고 유형</td><td>교통사고 외</td></tr>
<tr><td>사고 일시</td><td>2016년, 2017년</td><td>사고 지속시간</td><td>1년 이상</td></tr>
<tr><td>사고 내용</td><td colspan="4">
1. AR
AR(증강현실)은 가상현실(VR)의 한 분야로 실제로 존재하는 환경에 가상의 사물이나 정보를 합성하여 마치 원래의 환경에 존재하는 사물처럼 보이도록 하는 컴퓨터 그래픽 기법으로 디지털 미디어에서 빈번하게 사용된다.

2. 포켓몬고(Pokemon Go)
위치 기반 AR 게임 ‘포켓몬 고(Pokemon Go)’는 포켓몬이 실제 세계에 존재하지 않지만 현실 세계와 영상 콘텐츠의 결합으로 인해 이용자들이 포켓몬스터를 잡아 볼 수 있는 경험을 제공한다. 포켓몬 고를 실행한 후에 밖을 나가서 카메라를 비추다 보면 포켓몬을 만날 수 있고, AR 기반으로 현실 세계에 나타난 포켓몬을 볼 수도 있다. 포켓몬 고는 전 세계 사람들이 일반적으로 다 알고 있는 포켓몬스터 라는 콘텐츠와 AR의 결합으로 뜨거운 바람을 일으켰다.

3. 관련 교통사고
1) 국내
2017년 2월 7일 오후 10시 10분 정도에 승용차를 운전하던 운전자가 횡단보도를 건너던 보행자를 들이받았다. 이 사고로 보행자는 오른쪽 다리에 타박상을 입는 등 전치 2주의 부상을 입었다.
운전자는 운전을 하던 중 진행 방향 좌측에 포켓몬이 나타나자 이를 잡기 위해 급하게 좌회전을 하다가 횡단보도롤 건너던 보행자를 보지 못하고 사고를 냈다. 이 사고는 국내에서 포켓몬고로 인해 실제 교통사고가 발생된 첫 번째 사고이다.

2) 일본
① 2016년 8월 39세의 남성 운전자가 운전 중 포켓몬 고 게임을 하다가 행인 두 명을 치었고 그 중 한 명은 사망했다. 운전자는 두 명의 보행자를 보지 못
</td></tr>
</table>

	했다고 진술했으며, 이 사고는 일본에서 발생한 포켓몬 고 관련 첫 사망사고이다, ② 2016년 10월 횡단보도를 건너던 한 초등학생이 포켓몬 고를 하면서 운전을 하던 운전자의 트럭에 치여 사망했다. 이에 유족들은 관련 사고의 단속과 처벌을 강화하도록 국가에 요청했고, 포켓몬 고를 운영하는 나이앤틱은 일정 속도 이상을 감지하면 게임 속 아이템을 얻을 수 있는 가상의 장소인 '포켓스탑(poke-stop)'이 작동하지 않도록 2016년 11월에 설정을 바꿨다. 3) 미국 미국 인디애나주 퍼듀대학교 연구진들은 2017년 11월 21일(현지시각) '포켓몬고로 인한 죽음(Death by Poke'mon Go)'이라는 제목의 보고서를 발표하고 '포켓몬고' 때문에 미국 내 교통사고의 부상 및 사망률이 현저하게 증가했다고 밝혔다. 보고서에 따르면 연구진들은 2015년 3월부터 2016년 11월까지 인디애나주 티페카누 카운티에서 발생한 1만2000여건의 교통사고 데이터를 분석했다. 그 결과 '포켓몬 고'가 출시된 2016년 7월 이후 교통사고 발생률이 크게 증가했으며, 특히 몬스터볼을 획득할 수 있는 '포켓스탑'의 100미터 이내 지역에서는 교통사고가 기존에 비해 26.5% 증가했다는 사실을 알아냈다. 4) 교통사고에 대한 포켓몬 고 운영사 대처 포켓몬 고를 운영하는 나이언틱은 운전 중에 '포켓몬고'를 할 수 없도록 다양한 예방책을 마련했다. 걷는 속도보다 빠르게 움직이면 "운전중에 포켓몬고를 하지말라"는 메시지가 팝업되며, 움직이는 차량에 탑승했을 때는 운전자가 아닌 단순 승객인지 주기적으로 확인하는 과정을 거친다. 또 시속 30마일이 넘으면 포켓몬스터가 등장하지 않는다. **4. 위험지역 혹은 출입 급지 구역 침입 사례** 1) 2016년 6월 미국 오하이오 주에서는 포켓몬 고 게임을 하던 10대 청소년 3명이 원자력발전소 부지로 들어갔다가 경비원들에 의해 쫓겨났다. 2) 보스니아에서는 포켓몬 고 이용자들이 지뢰밭으로 들어간 사례가 보고되었다. 3) 도쿄 미나토(港)구의 한 지하철역에서는 전차에 포켓몬이 출현해 사고가 날 뻔한 일도 있었다. 4) 국내에서는 부산 유엔기념공원에 출입이 금지된 야간에 담을 넘어 내부로 들어오는 사람들이 나타났다. 부산시민공원 측은 공원 내 조명을 모두 끄고 출입을 통제하는 자정 이후에도 포켓몬 고를 하려는 사람들이 있어 출입 통제에 어려움을 겪고 있다고 밝혔다. **5. 위와 같은 사례에 대한 운영사 대처** 나이앤틱은 원자력발전소와 군부대 등 위험한 장소와 사유지 등은 앞으로 당사자가 신청하면 포켓몬 고 게임 장소에서 제외가 가능하다고 밝혔다.

사고 피해 내역	**1. 피해 내역:** 1) 보행 중, 혹은 운전 중 포켓몬 고 게임을 하는 유저들에 의해 최대 사망에 이르는 교통사고 다수 발생 2) 철도의 선로 위, 원자력 발전소의 일반인 출입 금지 지역, 재해 지역의 피난 명령 지역 등에서도 포켓몬이 출현해 이용자들의 침입으로 인한 안전사고 우려

SW신기술 안전사고 보고서

No. Metaverse-07

<table>
<tr><td>사고 분야</td><td colspan="4">VR/Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">메타버스의 플랫폼화를 가로막는 주요 난제 중 하나인 VR 울렁증</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://m.hankookilbo.com/News/Read/A2022051114510002989</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022.05.17</td></tr>
<tr><td>작성자</td><td colspan="3">이승엽</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td></td><td>사고 유형</td><td>VR 울렁증</td></tr>
<tr><td>사고 일시</td><td></td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. VR 울렁증
가상현실(VR)에 대한 진입 장벽이 점점 낮아지면서 저렴한 독립형 헤드셋을 구매해 VR 게임을 즐기는 등 가정에서도 VR을 쉽게 경험할 수 있다. 이렇게 VR의 접근성은 좋아졌지만, 메타버스의 핵심기술인 가상·증강현실(VR·AR) 헤드셋을 사용할 경우 사람에 따라 멀미를 심하게 느끼는 경우가 종종 발생하고 있다.
게임업계에 따르면 VR 울렁증은 3D, VR 게임 이용이나 영상 시청 시, 일부 이용자들이 겪는 멀미와 비슷한 육체적 증상을 뜻한다. 보통 1인칭 시점으로 진행되는 콘텐츠에서 많이 발생하는데 피로감이나 두통, 구토 등을 동반하기도 한다.
평면으로 된 모니터 화면보다 인간의 시야를 전부 감싼 VR 화면에서 훨씬 심한 멀미 증상을 보이기 때문에 1990년대 3D 게임에서 처음으로 등장했던 멀미 현상은 VR기기 보급과 더불어 대중에 더 널리 알려졌으며 영미권에서는 VR 울렁증을‘VR Sickness’로 지칭할 정도다.

2. 멀미와 VR 울렁증
1) 멀미
원인과 증상은 일반적으로 차량이나 배에 탑승했을 때 생기는 멀미와 동일하다. 멀미의 원인이 밝혀진 것은 아니지만 일반적으로 가장 설득력 있게 받아들여지는 이론은 ‘감각충돌론'이다. 보통 인간은 눈에 들어오는 시각 정보와 귓속의 반고리관 및 전정기관의 평형감각을 통해 신체의 평형감각까지 조절한다. 그런데 시각 정보와 평형감각에 '불일치'가 발생할 경우 뇌가 피로감을 느끼고 이를 독성물질에 의한 중독 등으로 판단, 구토 등을 통해 신체를 보호하려는 게 바로 멀미다.
2) 멀미와 VR 울렁증의 유사점
멀미가 버스에 탔을 때 신체는 가만히 있지만 주변 시야는 계속해서 이동해</td></tr>
</table>

괴리가 발생할 때 나타난다면, 가상 세계 속 멀미는 캐릭터가 괴물을 피해 달리고 총을 쏘는 등 격렬하게 움직이지만 현실 세계의 이용자는 의자에 앉아서 손만 움직이기 때문에 나타난다. 눈은 '사용자가 움직이고 있다'고 뇌에 알려주지만 귓속의 평형기관은 '사용자가 움직이지 않고 있다'고 알려주기 때문이다.

3) VR 울렁증의 원인

VR 울렁증은 기술적인 이유에서 일반적인 멀미보다 증상이 심한 것으로 알려졌다. 주요 원인 중 하나는 지연(레이턴시) 문제다. 레이턴시는 입력장치와 출력장치 사이에 전기 신호가 전달되는 시간을 뜻한다. VR 헤드셋을 예로 들면 사용자가 고개를 우측으로 돌렸을 때 즉시 디스플레이도 화면을 오른쪽으로 전환해야 감각의 괴리를 줄일 수 있다. 하지만 레이턴시가 길어지면 실제 움직임과 시각 정보의 차이가 생겨 멀미 현상이 발생할 확률이 높다.

대부분의 VR 게임이 사용하는 1인칭 시점도 VR 울렁증 발생에 큰 영향을 미치는 것으로 전해졌다. 보통 캐릭터의 등을 바라보며 진행하는 3인칭 시점의 게임보다 이용자 본인이 직접 세상을 바라보고 진행하는 1인칭 시점 게임에서 멀미가 자주 발생한다. 특히 최신 게임의 경우 시각적 효과를 위해 자주 시야를 흔들거나 회전하는데 이로 인해 뇌의 혼란이 가중된다. VR 게임의 장점인 '현실과 같은 몰입감'이 오히려 VR 게임을 즐기는 데 장애물이 된 것이다.

3. 해결책

보통 멀미의 근본적인 해결책은 두 감각기관의 정보를 일치시키는 것이다. 멀미 해소법으로 버스에 탔을 때 흔들림이 적은 앞 좌석에 앉거나, 창문으로 먼 산을 바라보라고 하는 것도 이 때문이다. VR 울렁증이 이슈가 됐던 초기에도 멀미약을 복용하거나 긴 시간 게임 이용을 자제하고 휴식을 취하는 등의 해결법이 제시됐다. 하지만 VR 기술이 고도화되면서 VR 울렁증을 기술적으로 해결하려는 움직임도 잇따르고 있다.

1) 일본

일본 소니는 세계지식재산권 기구에 VR 멀미 방지기술 특허를 공개했다. VR 헤드셋에 부착된 장치가 착용자의 머리에 진동을 가해 멀미를 최소화하는 것이 핵심이다.

2) 국내

한국전자통신연구원은 지난 2020년 VR 콘텐츠 편집 도구를 개발했다. 게임 등 콘텐츠 개발자가 개발 단계에서 멀미에 영향을 미치는 시점 높낮이, 배경 그래픽, 카메라 이동 속도 등을 세세하게 조절해 멀미를 최소화할 수 있도록 한다.

VR기기 업체에서는 게임에 따라 '멀미' 등급을 나눠 이용자에게 편의를 제공

	하기도 한다. 오큘러스는 VR게임을 △편안함 △보통 △움직임 많음 등의 3가지 등급으로 나눠 이용자들에게 경고한다. 3) 레이턴시를 줄이고 뇌의 연산 속도를 따라잡을 수 있는 디스플레이와 착용형(웨어러블) 기기 등의 개발도 한창이다. 멀미를 억제하기 위해 실제와 같은 고화질 화면(해상도)을 높은 프레임(1초당 보여주는 화면 수)으로 제공하는 방식이다. VR업계에서는 VR 기기는 일반적으로 멀미 유발을 줄이기 위해 밝기 1,000니트(nt) 이상, 해상도 60PPD(Pixel Per Degree) 이상, 주사율 120헤르츠(Hz) 이상의 디스플레이를 지원해야 한다고 보고 있다. 트레드밀과 모션 컨트롤러, VR 슈트 등 몸을 움직이면서 게임이 가능한 장치도 상용화를 앞두고 있다.
사고 피해 내역	**1. 피해 대상:** 3D, VR 게임의 일부 이용자 **2. 피해 내역:** 3D, VR 게임 이용이나 영상 시청 시, 일부 이용자들이 멀미와 비슷한 피로감이나 두통, 구토 등을 동반한 육체적 증상을 겪음

SW신기술 안전사고 보고서

No. Metaverse-08

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">메타버스 속 집단 성범죄 피해</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.codingworldnews.com/news/articleView.html?idxno=7925</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 4월 6일</td></tr>
<tr><td>작성자</td><td colspan="3">이선영</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>메타버스 속 성범죄</td></tr>
<tr><td>사고 일시</td><td>2022년 1월 30일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 메타버스 플랫폼
사건이 발생한 가상세계의 이름은 ‘호라이즌 월드(Horizon Worlds)’이다. 호라이즌 월드는 기존 페이스북, 현재 메타에서 개발한 메타버스 플랫폼이다. 메타는 2014년 헤드셋 전문업체 오큘러스(Oculus)를 인수했으며 메타버스 플랫폼 구축에 대한 개발에 꾸준히 투자해왔다. 그 결과 2021년 12월 9일, 메타버스 기반의 가상현실 공간 호라이즌 월드를 정식 출시했다.
호라이즌 월드는 이용 가능한 국가 내 18세 이상의 이용자들만 참여할 수 있다. VR 기기를 통해 접속할 수 있으며 엄지손가락을 치켜들어 상대방에게 웃는 표정을 노출하거나 엄지손가락을 내려 우는 표정을 노출하는 듯 다양한 형태로 의사소통을 진행할 수 있다. 새로운 사람과의 교류나 차단이 자유롭게 가능하며 상반신이 구현되어 있는 아바타와 이용자가 직접 설계하고 제작할 수 있는 가상 공간도 존재한다. 메타버스의 저성장 흐름 속 메타버스를 선도할 수 있을 것이라는 긍정적인 평가와 아직은 다소 부족하다는 부정적인 평가들이 공존하고 있는 상태이다.

2. 사고 경위
1) 보도 내용
영국 여성 ‘니나 제인 파텔(Nina Jane Patel)’이 메타버스 속에서 집단 성범죄를 당했다고 밝혔다. 테크 기업 카부니 벤처스(Kabuni Ventures)의 메타버스 연구 부사장 니나 제인 파텔이라고 이름을 밝히고 공론화를 한 피해 여성은 “메타(Meta)의 가상 게임 호라이즌 월드(Horizon Worlds)에서 가입한지 60초 만에 남성 아바타와 남성 목소리를 한 다른 사용자 3~4명이 범한 집단 성범죄 피해를 당했다. 그들은 사실상 갱단과 같았다.”라고 작성했다. 이어서 “가해자는 나의 아바타 사진을 촬영하고 ‘성적 행위를 싫어하는 척 하지 마라’ 등과 같은 메시지를 보냈다”라고 덧붙였다. 그녀는 그것을 "너무 빨리 일어난 끔찍한 경험"이라고 묘사하여 "안전 장치”를 사용할 생각을 할 시간조차 없었다. 파텔은 현재 미국과 캐나다 18세 이상 사용자 누구나 호라</td></tr>
</table>

	이즌 월드에 접속할 수 있다는 점과 최근 들어 메타버스 성범죄 피해 사례가 여러 차례 보고된 점을 지적하며, 문제의 심각성을 강조했다. 2) 후속 조치 사이버 공간 및 SNS 조사 전문 기관 보스코 리걸 서비스(Bosco Legal Services) 회장 조셉 존스(Joseph Jones)는 파텔처럼 가상세계에서의 범죄 피해를 당한 후 자신의 사례를 밝히는 이들은 많지 않지만, 실제 메타버스가 성범죄 공간이 되는 사례가 급격히 증가하고 있다고 말했다. 존스 회장은 메타버스 속 범죄가 해당 사용자의 발언, 피해자 아바타의 개인 식별 정보 포함 여부 등을 포함해 가상세계 범죄 사건을 다뤄야 한다고 말했다. 그는 “그녀의 사건에서, 가해자들이 그녀 본인의 신원을 확인할 수 있는 정보가 없었기 때문에 성범죄 소송은 불가능하다. 하지만 이후 비슷한 문제가 발생하여 또 다른 피해자가 등장하지 않도록 민사 소송을 제기할 수 있게 하겠다.”라고 말했다. 이어서 “여기서 발생하는 문제는 가해자도 익명성을 지녔기 때문에 실질적으로 메타버스 내 범죄 가해자를 상대로 민사 소송을 제기하기 어렵다는 것이다”라고 말했다. 3) 메타버스 속 성범죄 사례 2022년 4월 14일, 한국 경기북부경찰청 사이버범죄수사대는 아동·청소년의 성보호에 관한 법률 위반 혐의를 받는 30대 남성을 구속했다. 메타버스 플랫폼 ‘제페토(ZEPETO)’에서 아동·청소년 11명에게 접근, 신체 부위 등을 촬영해 달라고 요구한 사건이다. 2021년 11월 26일, 한 여성이 메타의 호라이즌 월드의 베타 테스트 도중 한 남성 그룹에게 성희롱을 당했다고 신고를 했다. 그녀는 “일반 인터넷에서의 성희롱 장난을 넘어 VR을 사용하고 있기 때문에 더 격렬한 행위들을 당했다”라고 밝혔다. 이어서 “어젯밤 온 몸이 더듬어졌을 뿐만 아니라 주위에 그 행위를 지지하는 다수의 인원들 때문에 해당 공간에서 매우 고립된 느낌을 받았다”라고 발언하여 가상 환경 속 성범죄의 위험성을 알렸다. **3. 시사점** 메타버스를 정의하기는 여전히 어렵지만, 사람들은 현재 소셜 미디어 플랫폼의 AI 또는 경찰 중심 사건 해결을 통해 성폭력에 대한 역사적 해결책에 의존하고 있다. 메타버스에서 발생한 성폭행 사건들이 계속 발생하고 있고 이제는 익숙해져 버린 수준이다. 메타버스라는 기술이 새로운 것이기는 하지만, 성폭력의 위협은 우리가 물리적 세계와 온라인 세계 모두에서 익숙한 심각한 문제의 연속이다. 메타버스 플랫폼을 운영하는 회사들은 지금까지 다양한 문제에 대해 사용자를 실망시켰기 때문에 메타버스가 우리의 일상에 포함되기 전에 회사들에 더 많은 책임과 명확한 해결책을 요구하는 것이 중요하다. 하지만 우리는 어떤 매체에서도 성폭행 문제에 대한 마법적인 기술적인 해결책이 없다는 사실을 받아들여야 할 것이다. 우리는 메타버스에서 성폭력이 말끔하게 고쳐지거나 조작될 수 있는 문제가 아니라는 것을 인정하고 다

	양한 규제와 회사들의 대처로 최대한 예방하고 처벌해야 할 필요가 있다.
사고 피해 내역(예상)	1. **피해 대상**: 영국 여성 니나 제인 파텔 2. **피해 내역**: 메타버스 속 다수의 남성들에게 집단 성희롱을 당한 사건

SW신기술 안전사고 보고서

No. Metaverse-09

<table>
<tr><td>사고 분야</td><td colspan="4">Metaverse</td></tr>
<tr><td>사고명</td><td colspan="4">트렌드마이크로, 메타버스 악용의 공간 ‘다크버스’경고</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.boannews.com/media/view.asp?idx=109374</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 8월 26일</td></tr>
<tr><td>작성자</td><td colspan="3">김영명</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>메타버스 악용 경고</td></tr>
<tr><td>사고 일시</td><td>2022월 8월 26일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 관련 용어
메타버스는 빠르게 발전하고 있다. 하지만 메타버스가 크게 성장함에 따라 이를 악용하며 본인의 이득 혹은 다른 이들의 피해를 주기 위한 다양한 위협 행위들도 끊임없이 발전하고 있다. 미국의 글로벌 보안 소프트웨어 기업 트렌드마이크로(TrendMicro)가 ‘인터넷 경험을 노리는 사이버보안 위협’ 보고서를 발표했으며 여기서 메타버스(Metaverse)의 다크웹 버전이라고 불리는 ‘다크버스(Darkverse)’가 등장했다.
다크버스는 모두가 이용하는 사이버공간의 허점을 파고들어 본인들만의 이익을 취하는 악성 공간으써 활용되고 있다. 다크버스는 메타버스 내부에 존재한다는 점을 제외하면 다크웹과 비슷하다. 다른 방면에서 다크버스를 봤을 때 평범한 사용자는 사이버상의 존재와 물리적 존재에게 동시에 영향을 받는 사이버-물리 때문에 다크웹보다 더 위험하다. 사이버-물리란 가상세계와 물리적 실체가 연동되는 것으로, 가상의 영역에 속하는 컴퓨팅, 통신, 제어를 실제 물리적 세계와의 통합을 의미한다. 다크웹 범죄 포럼의 온라인 공개 토론의 장에서는 다양한 위협들에 대해 끊임없이 논의되고 있다.

2. 다크버스의 위협 시나리오
1) 과도한 자유
다크버스는 불법 또는 범죄 활동을 조장하고 수행하기 위해 만들어졌지만, 오히려 이 공간은 억압적인 단체나 정부에 대한 언론의 자유를 위해 사용될 수도 있다. 이는 불법적인 단체 형성에 이바지할 수 있다는 문제점을 가지고 있다.

2) 불법 거래
다크버스는 메타버스의 암거래를 위한 지하 시장이 될 수 있다. 다크버스에서 만들어진 시장은 불법 또는 범죄 활동에 사용되며 사용자는 액세스를 위해 거액 혹은 소액의 인증 토큰이 필요하게 된다. 불법 거래를 막기 위해 만들어진 법 집행기관에서 이러한 공간을 사전에 알고 있더라도 시장에 접근하</td></tr>
</table>

	기 위한 인증토큰이 없이는 침투와 조사가 불가능하다. 다크버스 내 사용자 공간은 눈에 잘 띄면서도 법의 통제를 받기는 쉽지 않다는 것을 의미한다. 3) 커뮤니티 과보호 다크버스 사용자는 지정된 특정한 물리적인 위치 안에 있는 경우에만 다크버스 공간에 액세스할 수 있다. 이것은 메타버스에서의 폐쇄적인 커뮤니티에 대한 높은 등급의 보호를 받게 된다. 4) 로그 기록 불가 다크버스 사용자는 위치 기반 메시지와 메타버스 공간에 대한 근접성 메시지를 구현할 수 있으며, 이것은 법 집행기관에서 사용자가 서로 교환하는 통신을 가로채는 것을 극도로 어렵게 만들 수 있다. 그로 인해 이용자에 대한 통신 기록을 남기는 것에 큰 어려움이 생길 수 있다. 5) 범죄 추적 어려움 법 집행기관은 다크버스에서의 아동 성행위나 성폭행을 감시할 방법이 없기 때문에 관련된 성범죄는 증가하겠지만, 법 위반자를 추적하기가 매우 어려워진다. 그에 따라 범죄를 예방하는 것은 물론 범죄자를 추적하는 것에 큰 문제가 발생한다. **3. 시사점** 다크버스는 메타버스 형식의 다크웹과 유사한 형태로 이용되고 발전될 것이므로, 여러 범죄자들이 불법행위를 공모하고 실제로 실행에 옮겨도 처벌받지 않을 수 있다. 이용자는 물리적으로 지정된 장소에 있을 경우에만 다크버스에 접속이 가능하기 때문에 기존 폐쇄형 범죄 커뮤니티에 추가적 보호장치가 있는 셈이다. 따라서 경찰도 정확한 인증 토큰 없이는 다크버스에서 운영되는 지하시장에 잠입하는 것이 불가능에 가까울 것으로 예상된다. 이러한 특성에 따라 다크버스를 은신처로 삼으며 금융·전자상거래 사기부터 NFT 절도, 랜섬웨어 등에 이르기까지 다양한 보안 위협이 증가되고 사건사고들이 발생될 전망이다. 또한, 메타버스의 사이버 물리 특성이 위협 행위자들에게 새로운 불법행위의 활로를 열어줄 것으로 보인다. 인프라 관리자들이 운영하는 '디지털 트윈' 공간을 위협해 산업 시스템을 파괴하거나 왜곡하는 사이버 범죄 행위도 발생할 것으로 예상된다. 이미 실제 사례로 메타버스 이용자들의 아바타에 멀웨어를 유포해 물리적 손상을 일으키는 범죄행위는 이미 여러 차례 보고되기도 했다. 메타버스와 같은 사이버 공간이 빠르게 보편화될 것으로 분석되는 가운데 어떤 사이버위협이 예상되고 어떻게 완화할 수 있는지에 다음과 같은 질문으로 논의가 필요할 것이다.
사고 피해 내역(예상)	**1. 피해 대상:** 불특정 다수 **2. 피해 내역:** 금융·전자상거래 사기부터 NFT 절도, 랜섬웨어 등에 이르기까지 다양한 보안 위협과 사건사고에 대한 경고

SW신기술 안전사고 보고서

No. Blockchain-01

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">크로스 체인 브릿지에서 발생한 대규모 해킹</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.fnnews.com/news/202203311612425421</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 3월 31일</td></tr>
<tr><td>작성자</td><td colspan="3">정영일</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>액시 인피니티</td><td>사고 유형</td><td>해킹</td></tr>
<tr><td>사고 일시</td><td>2022년 3월 23일</td><td>사고 지속시간</td><td>일주일</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 플랫폼
‘액시 인피니티(Axie Infinit)’는 2018년에 출시되었으며, 이른바 돈버는 게임 P2E(Play to Earn) 트렌드를 대표하는 게임이다. '액시 인피니티'의 플레이어들은 '엑시즈(Axies)'라 불리는 대체불가능토큰(NFT) 형태의 디지털 애완동물을 수집해 다른 액시즈와 싸우고 양육한다. 이 액시즈를 팔면 암호화폐를 벌 수 있으며, 특히 필리핀에서 그 인기가 높아 온종일 게임을 하는 이용자가 늘어나면서 해당 게임의 플레이어가 수익성 높은 직업으로 떠오르기도 했다.

2. 관련 블록체인
액시 인피니티의 개발사인 베트남의 '스카이마비스'가 소유한 블록체인 네트워크인 '로닌 네트워크'를 통해 플레이어들은 자신들이 액시 인피니티에서 번 디지털 코인을 이더리움 같은 다른 암호화폐와 교환할 수 있다.
로닌 네트워크는 스카이마비스가 엑시 인피니티 구동을 위해 지난해 2월 자체 개발한 이더리움 사이드체인으로 액시 인피니티가 이더리움 메인넷을 사용하며 문제가 돼 왔던 높은 가스비(거래 수수료)를 낮추기 위해 개발됐다. 2021년 4월 28일 이더리움에서 로닌 네트워크로의 이전을 모두 마친 이후부터 액시 인피니티는 로닌 네트워크 기반으로 가동되고 있다.
액시 인피니티와 이더리움 등 다른 블록체인을 연결해주는 이른바 ‘로닌 브릿지’를 통해 이용자는 이더리움이나 스테이블 코인 등을 예치한 뒤 게임용 통화나 NFT를 구매해 게임을 할 수 있고, 게임이 끝나면 이를 다시 팔아 돈을 인출 할 수 있다.

3. 사고 경위
1) 사고인지
액시 인피니티의 한 이용자가 자신이 보유한 5천 이더리움을 인출 할 수 없다고 신고하였고, 이를 조사하는 과정에서 해킹 사실이 뒤늦게 밝혀졌다. 사고는 2022년 3월 23일 발생했지만 액시 인피니티 개발사인 스카이마비스는</td></tr>
</table>

엿새 뒤인 2022년 3월 29일 사용자들이 이더리움 출금이 안된다는 신고를 한 후에야 인지하고 파악에 나선 것으로 알려져 더욱 논란이 되었다.

2) 해킹 방법
로닌 네트워크가 공개한 사고 보고서에 따르면 이번 해킹은 검증자(validator) 노드의 프라이빗키(private key)를 해킹하는 방식으로 이뤄졌다. 로닌 네트워크는 총 9개의 검증자 노드로 구성돼 입출금이 발생할 때마다 이들 중 5개 이상의 노드의 승인을 받게 돼 있는데, 해커가 5개 노드의 프라이빗키를 해킹하는 데 성공한 것이다. 해킹을 당한 노드는 스카이마비스 노드 4개와 엑시 다오(DAO, 탈중앙화자율조직)에서 운영하는 제 3자 노드 1개다. 해커들은 로닌 브릿지에서 수 억 달러 가치의 암호화폐를 가져갔다.

3) 사고 발생지점
크로스체인 브릿지란 서로 다른 블록체인 플랫폼 간 자산 이동을 돕는 기술로, 탈중앙화 금융(디파이) 시장에선 필수적이다. 로닌은 액시 인피니티가 게임 속도 향상을 위해 개발한 사이드체인으로 이번 사건은 액시 인피니티와 이더리움 등 다른 블록체인을 연결해주는 이른바 '로닌 브릿지'에서 발생했다.

4) 시사점
이번 해킹이 복수의 블록체인을 연결 시켜주는 '브릿지'를 노렸다는 점에서 유사한 기술을 사용하고 있는 여러 디파이(DeFi·탈중앙화금융)의 신뢰성까지 흔들고 있다는 분석이 나오고 있으며, 로닌 블록체인에 참여하고 있는 검증자가 충분히 많지 않아 발생한 문제라는 지적도 나왔다.

4. 해킹에 대한 입장 및 조치
로닌 네트워크 측은 "범죄자들을 처벌하기 위해 다양한 정부 기관들과 직접 협업하고 있고 로닌 네트워크에 있는 액시 인피니티(AXS), 로닌(RON), 스무스러브포션(SLP)은 모두 안전하다"고 설명했다. 그리고 3개월 동안 로닛 브릿지 작동을 중지하였다.

5. 결과
2022년 6월 28일(현지시간) 액시 인피니티 운영사 스카이마비스는 해킹 피해 발생 후 가동 중지했던 로닛 브릿지 가동을 재개했다고 알리며 해킹으로 탈취된 이용자 소유 암호화폐에 대해선 전액 변상을 완료했다고 알렸다. 이날 로닌 브릿지 공식 블로그에 올라온 공지에 따르면 모든 이용자 소유 자산은 새로운 브릿지에서 1:1로 백업됐다. 액시 탈중앙화자율조직(DAO·다오) 소유의 11만 7600 ETH와 2550만 USDC, 해킹 사건 이전에 이미 바이낸스 거래소가 운영하는 브릿지로 옮겨간 4만 6000 ETH를 제외한 총 11만 7600 ETH와 2550만 USDC가 이용자 피해 변상에 쓰인다.
피해 보상금은 외부 자금을 조달해 마련했다. 액시 인피니티는 지난 4월 로닌

	브릿지 해킹으로 유출된 자금을 변상하는 데 필요한 자금을 모은다며 1억 5000만 달러 규모의 투자를 유치한 바 있는데 바이낸스와 a16z 등 대형 암호화폐 벤처 투자사(VC)들이 투자에 참여하면서 관심을 모았다. 한편, 액시 인피니티 해킹 사건 배후로는 북한의 해킹 그룹 '라자루스(Lazarus)'가 지목되었다.
사고 피해 내역	1. **피해대상**: 로닌 브릿지를 통해 가상화폐를 인출하고자 한 액시 인피니티 이용자 2. **피해내역**: 해킹으로 인해 이더리움 코인 17만3600개와 USDC 2550만개로, 한화 기준 7400억~7600억원의 재산상의 피해를 입었으며, 이 피해는 해킹으로 역대 두 번째로 큰 규모임

SW신기술 안전사고 보고서

No. Blockchain-02

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">알고리즘 기반 스테이블 코인의 폭락으로 인한 투자자 손실</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hani.co.kr/arti/economy/economy_general/1043744.html</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 5월 21일</td></tr>
<tr><td>작성자</td><td colspan="3">박근모</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>가상자산 시장</td><td>사고 유형</td><td>가상화폐 가치 폭락</td></tr>
<tr><td>사고 일시</td><td>2022년 5월</td><td>사고 지속시간</td><td>약 1주일</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 플랫폼
‘한국판 머스크’로 불리는 권도형 테라폼랩스 대표가 만든 테라는 금융 서비스에 특화된 블록체인으로 이름을 알렸다.
가상화폐는 하루에 수십 %씩 오르내리는 변동성이 취약점으로 꼽혀왔는데, 테라는 달러 등 실물자산에 연동하도록 설계된 스테이블 코인 방식을 활용해 가격 변동성을 줄인다는 장점을 내세웠다.

2. 테라의 작동원리
1) 스테이블 코인
스테이블 코인은 말 그대로 1코인의 가치를 ‘일정하게’(stable) 유지하는 코인을 말한다. 대다수의 스테이블 코인은 미국 달러화와 연동해 ‘1코인=1달러’ 가치를 유지하도록 설계한다. 비트코인, 이더리움 등 일반적인 코인은 짧은 시간에 가격 변동성이 심해 결제나 디파이(DeFi·탈중앙화금융)에 사용하기 어렵다. 하지만 스테이블 코인은 가치가 항상 일정하다 보니 이런 점에서 유리하다.
스테이블 코인은 크게 담보물 기반(미국 달러 등 법정화폐나 또 다른 가상자산을 담보로 가치를 일정하게 유지)과 알고리즘 기반으로 구분하는데, 테라유에스디는 알고리즘 기반 스테이블 코인에 해당한다.

2) 작동원리
알고리즘 기반 스테이블 코인은 담보물이 필요 없는 대신 정교하게 짜인 프로그래밍 코드로 코인의 가치를 일정하게 유지하도록 설계한다.
테라유에스디는 알고리즘으로 가치를 유지하기 위해 루나라는 짝꿍이 필요하다. 평상시 1개의 테라유에스디는 1달러를 유지하겠지만, 만약 1달러 이하로 내려간다면 알고리즘은 테라유에스디의 가격을 다시 1달러까지 올리기 위해 루나를 새로 발행한다. 발행한 루나로 시장에 풀린 테라유에스디를 구입한다.</td></tr>
</table>

끝으로 앞서 구입한 테라유에스디를 소각한다. 이 과정을 반복하면, 테라유에스디 유통량이 줄어들어 1달러까지 오른다.
반대로 테라유에스디가 1달러 이상일 경우에는 테라유에스디를 새로 발행해 시장에 풀린 루나를 구입하고, 구입한 루나를 소각하는 작업을 반복한다. 결과적으로 1달러를 중심으로 테라유에스디와 루나의 유통량이 늘어나고 줄어들면서 가격을 유지한다. 그리고 이 모든 과정은 프로그래밍(알고리즘)이 돼 자동으로 이뤄진다.

3. 사고 경위

1) 루나는 2022년 5월 초만 하더라도 시가총액 300억달러(약 38조원)를 기록하며 한때 전 세계 가상자산 중 8위까지 올랐다. 당시 루나보다 더 높은 시가총액의 코인은 비트코인, 이더리움 등 7개에 불과했던 루나는 시가총액이 불과 열흘 만에 4900만달러(약 620억원)까지 폭락했고, 1루나당 117달러를 자랑하던 루나코인의 개당 가격은 5월 19일 기준 0.00014달러로 내려앉으면서 한순간에 99.99%까지 가격이 하락한 이번 루나·테라 파장이 국내를 넘어 전 세계 가상자산 시장에 악재로 작용했다.

2) 폭락 원인: 알고리즘의 헛점

이번 루나 폭락 사태는 테라유에스디가 1달러 유지에 일시적으로 실패하면서 발생했다. 알고리즘은 1달러 유지를 위해 루나를 발행하고, 발행한 루나로 테라유에스디를 구입하고, 테라유에스디 소각을 진행했다. 이 과정에서 루나의 유통량이 일시적으로 늘면서 가격이 하락했다. 당연한 일이었고, 이 정도는 알고리즘의 예상 범위였다.
하지만 전통 금융시장에서 금리 인상에 따른 시장 위축이라는 점은 미처 알고리즘이 고려하지 못하여 거시경제의 위축으로 비트코인과 이더리움이 하락하자, 시장은 한순간에 얼어붙었다. 엎친 데 덮친 격으로 순간적으로 루나의 유통량이 늘어나자, 루나의 낙폭은 더 크게 다가왔다. 알고리즘의 예상 범위를 벗어났다. 루나 가격이 하락하자, 테라유에스디를 구입하기 위해 더 많은 루나가 필요했다. 외부 상황을 알지 못하는 알고리즘은 루나 가격이 더 하락한다는 것도 모른 채 테라유에스디 가격 상승을 위해 루나를 더 발행할 뿐이었다.

3) 폭락에 대한 개발사 입장

권도형 테라폼랩스 대표는 인터뷰를 통해 전 세계 가상자산 시장을 얼어붙게 한 '루나 사태'에 대해 책임을 통감하고 이를 재건할 자신이 있다고 인터뷰를 통해 밝혔다.

4. 진행 상황

2022년 6월 12일 루나와 테라 폭락 사태를 수사하고 있는 한국 검찰이 정확한 피해 규모를 추산하는 작업에 나선 것으로 알려졌다. 피해액을 산정하려면

	개발사인 테라폼랩스로부터 관련 자료를 확보해야 하기 때문에 검찰이 강제수사에 나설 수도 있다는 전망이 나온다. 특히 검찰은 거래소에서 루나 코인을 구매했다 손실을 본 사례에도 사기 혐의를 적용할 수 있을지 법리 검토에 들어간 것으로 알려졌다.
사고 피해 내역	1. **피해대상**: 스테이블코인 루나 투자자 2. **피해내역**: 루나가 1주일 새 99% 폭락하면서 루나 시가총액이 52조 7000억원에서 이날 3조 8000억 원으로 고꾸라졌다는 점을 고려하면 투자자 피해액은 50조원을 넘을 것으로 추정됨

SW신기술 안전사고 보고서

No. Blockchain-03

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">NFT 거래소 해킹</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.sedaily.com/NewsView/26287C212G</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 2월 21일</td></tr>
<tr><td>작성자</td><td colspan="3">정혜진</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>오픈씨(OpenSea)</td><td>사고 유형</td><td>해킹</td></tr>
<tr><td>사고 일시</td><td>2022년 2월 19일</td><td>사고 지속시간</td><td>약 3시간</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 플랫폼
오픈씨(OpenSea)는 현재 세계 최대의 종합 NFT(대체 불가능한 토큰) 거래 플랫폼으로 사용자는 플랫폼에서 NFT를 발행, 전시, 거래 및 경매를 할 수 있다. 예술, 음악, 도메인네임, 가상세계, 트레이딩 카드, 스포츠 수집품, 게임 아이템 등 다양한 유형의 NFT를 사고 팔 수 있을 뿐만 아니라 오픈마켓에서 개인 창작물이 어떻게 사고 팔리는지를 볼 수 있다.

2. 오픈씨의 작동원리
NFT란 블록체인 암호화 기술을 활용해 JPG 파일이나 동영상 등 콘텐츠에 고유한 표식을 부여하는 신종 디지털 자산이며, 오픈씨는 스마트 계약 기술을 이용해 독특한 NFT를 사고 판매한다. 스마트 계약을 통해 사용자는 플랫폼에서 NFT 컬렉션을 안전하게 보호할 수 있고 크리에이터에게 보안을 제공한다. 플랫폼에서는 ERC721 및 ERC1155 표준을 사용하여 NFT 수집가가 시장에 나열된 자산을 소유하고 있는지 확인할 수 있다.

3. 관련 블록체인
오픈씨는 이더리움 블록체인을 사용한다. NFT를 사고 팔려면 메타마스크(MetaMask), 코인베이스(Coinbase), 트러스트월렛(TrustWallet) 등과 같은 이더리움 지갑이 필요하고, NFT를 구매하거나 판매하기 위해서는 사용자가 자신의 암호화폐 지갑에 연결해야 한다.
오픈씨는 이더리움을 비롯한 여러 블록체인을 기반으로 NFT 이동을 추적하며, 모든 거래는 암호화폐로 이루어진다. 오픈씨는 현재 이더리움 외에 솔라나(Solana), 클레이튼(Klaytn), 폴리곤(Polygon) 체인을 지원하고 있다.

3. 사고 경위
1) 사고 내용
2022년 2월 19일 오후 5시부터 8시까지 오픈씨 이용자 32명의 계정에서 비</td></tr>
</table>

	정상적인 NFT 전송 사고가 발생하여 일부 NFT 소유권이 무상으로 양도되었다. 데빈 핀저 오픈씨 최고경영자(CEO)는 “현재까지 총 32명의 사용자들이 해커의 악의적인 페이로드(스팸메일 등 악성 코드가 심어진 이메일)에 서명했으며, 일부 NFT가 도난당했다”고 밝혔다. 도난된 NFT 중에는 ‘지루한 원숭이 요트 클럽(Bored Ape)’, ‘아즈키’,‘두들스’ 등 유명 NFT가 포함된 것으로 알려졌으며 피해 규모는 170만 달러(약 20억원)를 넘을 것으로 보인다. 2) 범행 수법 이번 피싱은 NFT 스마트컨트랙트 표준처럼 쓰이는 '와이번 프로토콜'의 유연성을 악용한 것으로 보인다는 의견이 나왔다. 일단 많은 부분이 공백으로 남겨져 있는 컨트랙트에 이용자의 서명을 받은 후, 공격자가 마음대로 계약을 완성해 어떠한 대가 지불 없이 NFT 소유권을 이전한 것으로 보인다는 설명이다. 타깃이 된 이용자들은 백지수표에 서명을 한 셈이 되고, 공격자는 수표의 남은 빈칸을 채워 NFT를 가져간 것이다. **4. 오픈씨의 공식적, 대외적 입장 혹은 조치** 오픈씨는 어떤 공격 매개를 통해 서명에 이르게 됐는지 조사 중이라고 밝혔다. 다만 오픈씨 이메일이나 오픈씨가 진행하고 있는 컨트랙트 마이그레이션이 공격의 매개된 것은 아니라고 분명히 했고, 도난당한 일부 NFT는 회수됐다는 메시지를 남기며 이용자들을 안심시켰다. 오픈씨는 스마트 계약 표준을 새롭게 업데이트하고 있는 것으로 알려졌다. **5. 시사점** 오픈씨는 이번 사건은 '피싱'이라고 일축했지만, 오픈씨 플랫폼 자체를 노린 사이버 공격이 끊이지 않고 있다. 2022년 1월에는 버그를 이용해 고가 NFT를 약 10분의 1수준 가격에 탈취한 뒤 시세차익을 얻은 사례가 나왔고, 악성 NFT를 활용해 가상자산을 탈취할 수 있는 결함이 발견돼 오픈씨 측에서 긴급 점검에 나서기도 했다. 때문에 보안 업계에서는 NFT 광풍과 맞물려 유명 작가와 사이트, 직원 등을 사칭해 NFT를 탈취하는 사기 범죄를 비롯해 범죄 자금 세탁, 인위적으로 NFT 가치를 높이는 자전거래(Wash Trade) 등 유사 범죄가 잇따르고 있어 주의해야 한다고 당부하고 있다.
사고 피해 내역	**1. 피해 대상:** 오픈씨 사용자 32명 **2. 피해 내역:** 오픈씨 32명 사용자가 보유한 254개 NFT가 탈취되었으며, 170만달러(약 20억원) 상당의 피해 발생

SW신기술 안전사고 보고서

No. Blockchain-04

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">블록체인 기반 NFT 신종 사기</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://game.donga.com/102003/</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 1월 27일</td></tr>
<tr><td>작성자</td><td colspan="3">조학동</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>한국</td><td>사고 유형</td><td>러그풀</td></tr>
<tr><td>사고 일시</td><td>2021.11.25. ~</td><td>사고 지속시간</td><td>약 3개월</td></tr>
<tr><td>사고 내용</td><td colspan="4">
1. 사고 플랫폼
2021년 11월 25일 '캣슬' 이라는 명칭을 가진 NFT 프로젝트가 등장했다.
캣슬은 클레이튼 기반의 NFT로 탈중앙화 금융(디파이)을 하겠다는 로드맵을 통해 투자자(홀더)들을 유치했다. 희소 가치가 있는 고양이 캐릭터로 게임이나 상품을 만들어 투자 가치를 만들겠다는 사업 구상과 함께 매달 가상화폐를 지급하겠다는 약속도 함께 내걸었다.

2. 관련 블록체인
NFT란, 블록체인 기술을 이용해 그림이나 영상 등 디지털 자산에 소유권을 부여하는 방식으로 거래되는 신종 가상자산의 일종이다. 구매 즉시 소유권과 거래 내용이 블록체인에 저장되기 때문에 위변조할 수 없고, 그림 하나하나가 희소한 가치를 가지기 때문에 새로운 투자 수단으로 주목받고 있다.

3. 사고 경위
1) 등장 및 홍보
캣슬 프로젝트는 총 1만 마리의 각기 다른 고양이 NFT를 통해 탈중앙화 금융을 구축해 고양이를 구매하고, 이를 보유하면 킷(Kit)이라는 물고기를 준다고 홍보했다. 캣슬 NFT 10개를 가지고 있으면 매일 1킷(약 1클레이)을 주겠다고 하며 투자자 유치에 성공하였고 캣슬 프로젝트는 2021년 11월 출시 하루 만에 킷캣 1만 마리중 1차 물량인 1000개가 완판되는 성과를 보여주었다. 또한, NFT 거래 플랫폼인 오픈씨(opensea)에서 한때 거래량 2위, 클레이튼 NFT 마켓에선 6위까지 기록하였다.

2) 사기 수법 및 러그풀 발생
캣슬 프로젝트는 투자자들을 모집한 뒤 투자 가치가 있는 것처럼 보이기 위해 허위로 고양이 캐릭터를 거래하는 자전거래 수법을 동원하였다. 프로젝트 운영자 4명이 서로 고양이 캐릭터를 사고팔며 거래금액을 높였고, 최초 3만 원
</td></tr>
</table>

대에 팔리던 고양이 캐릭터는 최대 50만원까지 가격이 치솟았다. 이렇게 3백명이 넘는 투자자가 모집됐고, 2억 7천만원 상당의 고양이 캐릭터 5천개를 판매한 뒤 프로젝트 운영자들은 3개월도 안돼 잠적하였다.
킷캣 프로젝트의 운영자들은 2022년 1월 21일 “메인 계정 해킹으로 더 이상 프로젝트를 진행할 수 없다”는 말만 남기고 사라지고 관련 홈페이지와 오픈채팅, 사회관계망서비스(SNS) 등도 모두 폐쇄되었다.
러그풀이란 가상화폐 생태계에서 개발자가 갑자기 프로젝트를 중단하고 투자금을 들고 사라지는 사기 수법을 말한다. 양탄자(Rug)를 잡아당기면(Pull) 그 위에 있던 사람들이 한순간에 넘어진다는 비유적 표현에서 유래됐는데 이 캣슬 프로젝트 사건은 국내의 대표적인 러그풀 사례로 꼽히게 되었다. 캣슬이 러그풀을 한 1월 21일 이후 캣슬 NFT의 개당 가격은 3,000원 수준으로 폭락하였다.

3) 피해 신고
총 피해액이 50~60억 원에 이를 것으로 예상되고 있으며, 그 중 9명이 총 2억1천만 원 상당을 투자 사기를 당했다고 피해 신고를 하였다.

4) 검거
서울경찰청 사이버수사과는 2022년 4월 19일, 캣슬 NFT 개발자를 NFT를 이용해 2억 원대 사기 행각을 벌이고 도주한 혐의로 체포하였다. 또한, 범행을 도운 것으로 추정되는 4명도 사기 혐의로 불구속 입건해 공모 관계를 수사하고 있다고 밝혔다.

4. 증가하는 러그풀 사례
러그풀은 양탄자를 잡아 빼는 것처럼 가상화폐 시장에서 개발자가 프로젝트를 제대로 이행하지 않고 돌연 중단하면서 투자금을 가로채는 수법의 사기인데 개발자가 암호화폐와 디파이(DeFi)에서 갑자기 유동성을 회수하면서 결국 매도 악순환을 초래하여 투자자들에게 막대한 손실을 입힌다.
솔라나를 기반으로 하는 솔라이프 NFT는 관련 트위터와 홈페이지를 모두 닫고 사라졌다. 솔라이프는 NFT를 구해 돈 버는(P2E) 게임을 하는 프로젝트였다. NFT 뽑기 하나당 3.99솔라나(약 56만원)에 판매했는데, 현재는 가치가 폭락한 상태다. 국산 밈 코인으로 홍보한 떡볶이 코인(TBK)도 운영진이 잠적한 것으로 파악됐다. 사전 판매도 진행하고 NFT 까지 판매됐던 프로젝트다. 이 코인은 무려 이자율이 30%에 달하는 스테이킹(가상자산 예치·위임)으로 사람을 끌어모았지만, 운영진이 사라지고 공식 SNS 등도 모두 폐쇄된 상황이다.
이처럼 가상화폐와 관련된 범죄 피해액이 급증하고 있어 투자자들의 각별한 주의가 요구된다. 블록체인 데이터 플랫폼 기업 체이널리시스에 따르면 2021년 가상화폐 사기 범죄 피해액은 77억 달러(약 9조 2362억원)로 전년 대비 81% 증가했다. 특히 2020년 전체 가상자산 스캠(사기) 피해액의 1%에 불과했던 러그풀은 2021년 전체 스캠 피해 규모의 37%를 차지할 정도로 가파르

	게 증가하는 추세다.
사고 피해 내역	1. **피해대상**: 캣슬 프로젝트 투자자 2. **피해내역**: 300여명의 피해자, 피해금액이 약 50~60억원에 이를것으로 추정됨, 피해자 중 9명이 2억 1천만원 상당의 피해를 신고함

SW신기술 안전사고 보고서

No. Blockchain-05

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">솔라나 네트워크 장애</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">1)
https://www.fnnews.com/news/202109151704109109
2) https://www.news1.kr/articles/?4668071
3) https://www.etnews.com/20220602000196</td></tr>
<tr><td>보도 날짜</td><td colspan="3">1) 2021년 9월 15일
2) 2022년 5월 2일
3) 2022년 6월 3일</td></tr>
<tr><td>작성자</td><td colspan="3">1) 김소라
2) 김승준
3) 이형두</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>솔라나</td><td>사고 유형</td><td>블록 생성 오류</td></tr>
<tr><td>사고 일시</td><td>1) 2021년 9월 14일
2) 2022년 5월 1일
3) 2022년 6월 2일</td><td>사고 지속시간</td><td>1) 18시간
2) 7시간
3) 4시간 30분</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 사고 플랫폼
솔라나는 탈중앙화 분산 방식의 블록체인 시스템이 단일 노드 수준의 빠른 속도와 높은 성능을 구현하는 것을 목표로 하는 블록체인 플랫폼 또는 그 플랫폼에서 사용되는 암호화폐다.
솔라나는 NFT(Non-Fungible Token, 대체불가능한토큰)와 탈중앙금융(De-Fi, 디파이) 등 여러 블록체인 서비스들의 활용성을 높이면서도 이더리움보다 처리 속도가 빨라 인기를 모았다.

2. 사고 종류
1) 2021년 9월 15일
① 솔라나 블록체인에 초당 거래량이 40만 건이 몰렸고 이에 따라 솔라나 블록체인 노드 일부에 장애가 생겨 정상적인 블록 생성 및 거래 유효성 검증 역할을 수행하지 못하게 되면서 약 18시간 동안 네트워크 장애가 발생하였다.
② 사고 원인
업계는 블록체인에 과도한 거래량이 발생한 배경으로 솔라나의 디파이 프로젝트 레이디움(Raydium)을 지목했다. 레이디움에서 진행된 IDO(Initial DEX Offering, 탈중앙거래소 DEX에서의 토큰 공개)에서 봇의 압도적인 거래량으로 몇몇 노드에서 블록 생성이 중단됐고, 이에 따라 솔라나 블록체인 위에서 돌아가는 서비스들도 함께 멈춰버리는 사태가 빚어졌다는 설명이다.</td></tr>
</table>

2) 2022년 5월 1일

① 대규모 트랜잭션(초당 600만) 유입이 네트워크에 발생함에 따라 단일 노드에서 100Gbps를 초과하는 트래픽이 발생해 유효성 검사기의 메모리 부족과 충돌이 일어나 블록 생성이 중단됐으며 약 7시간 동안 지속되었다. 솔라나 네트워크를 이용하는 탈중앙화 분산 애플리케이션(DApp)의 일부 기능도 장애를 겪었다. 일부 애플리케이션 서비스는 솔라나 장애에 따른 송금 지연 등을 유의해달라는 공지를 이용자들에게 전했다.

② 조치

솔라나는 네트워크 안정성과 복구를 위해 현재 QUIC(구글이 구축한 트래젝션 제어에 사용할 수 있는 프로토콜), 지분 가중 트랜잭션 QoS(Stake-weighted transaction)와 수수료 기반 트랜잭션 실행 우선 순위(Fee-based execution priority) 지정, 솔라나 수수료 우선 순위 설정 등 세 가지 완화 조치를 진행 중으로 v1.11 릴리즈를 목표로 작업이 완료되는 대로 곧 배포하겠다는 입장을 밝혔다.

3) 2022년 6월 2일

① 2022년 5월 1일 이후 메인넷 중단 해결을 위한 완화조치를 추진하였지만 한 달 만에 솔라나 메인넷에서 새로운 블록 생성이 중단되면서 약 4시간 30분 가동이 정지됐다. 솔라나 네트워크에서 가동되는 디앱(dApp)들도 오프라인으로 전환됐다.

② 사고 원인

솔라나 측은 이번 사고의 원인을 블록체인의 '내구성 있는 논스 트랜잭션'(Durable Nonce Transaction)' 버그 때문이라고 설명했다. 논스는 암호학에서 특정한 목적을 위해 사용되는 난수를 의미한다. 해당 버그로 인해 노드들이 동일한 블록에 대해 서로 다른 결과를 생성했고, 이에 따라 최종 합의에 도달할 수 없게 돼 네트워크가 정지한 것으로 분석된다고 밝혔다.

③ 조치

솔라나는 “검증자들이 듀러블 트랜잭션 논스 기능을 비활성화한 후 네트워크를 재가동하기 위해 노력했다”며 향후 진행될 업데이트에서 오류가 수정될 것이라고 하였다.

4. 시사점

솔라나는 저렴한 거래 수수료와 빠른 전송 속도를 무기로 최근 급부상한 블록체인 플랫폼이다. 나스닥 거래소 운용을 염두에 두고 역사증명(PoH) 합의 알고리즘을 기반으로 설계돼 최대 71만TPS(초당거래건수) 속도가 가능하다고 주장하고 있으며, 주문을 신속하게 처리해야 하는 탈중앙화금융(디파이) 등 금융 섹터 위주로 급성장하며 시가총액 상위를 기록하고 있다. 하지만 솔라나는 낮은 수수료로 인해 오히려 분산서비스거부(DDoS) 공격이 잦아지면서 낮은 수수료가 양날의 검으로 작용한다는 평가를 받고 있다. 솔라나를 향한 디도스

	공격은 2021년에만 세 차례 이상 이뤄졌으며, 최근의 셧다운도 대체불가토큰(NFT) 발행 봇(Bot)들로 인해 발행량이 일시 급증한 것이 원인으로 분석됐다. 2021년 9월 장애 이후로는 자산 평가 플랫폼 쟁글 역시 '크립토 평가 보고서'에서 “솔라나는 검증자 노드 숫자가 늘고 있어 빠른 속도로 탈중앙화 수준이 개선되고 있지만 약 17시간 동안 셧다운된 사례가 있어 여전히 중앙화와 안정성 논란이 있다”고 평가한 바 있다.
사고 피해 내역	1. **피해 대상**: 솔라나 2. **피해 내역**: 블록 생성의 오류로, 암호화폐 거래 장애 및 해당 플랫폼을 기반으로 하는 서비스들의 네트워크 장애

SW신기술 안전사고 보고서

No. Blockchain-06

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">NFT 기반 게임 ‘엑시 인피니티’해킹</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.hankookilbo.com/News/Read/A2022041506080004474</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 4월 15일</td></tr>
<tr><td>작성자</td><td colspan="3">정상원</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>NFT 기반 게임 해킹</td></tr>
<tr><td>사고 일시</td><td>2022년 3월 29일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 게임 ‘엑시 인피니티’와 게임파이
‘엑시 인피니티(Axie Infinity)’는 베트남 스타트업 회사 스카이마비스(Sky Mavis)에서 개발한 수집형 모바일 게임이다. 엑시 인피니티 게임 내에선 ‘엑시’라는 가상의 동물을 모을 수 있는데 이 엑시를 암호화폐로 거래할 수 있다는 것이 특징이다.
최근 게임을 플레이하며 돈을 버는 PtE(Play to Earn) 게임이 많은 주목을 받게되면서 게임파이(GameFi) 산업이 본격적으로 성장하기 시작했다. 게임파이는 탈중앙화된 금융을 뜻하는 DeFi(Decentralized Finance)와 게임의 합성어다. 블록체인 기술과 NFT가 적극적으로 게임 시장에서 활용되면서 많이 활용되기 시작한 단어라고 볼 수 있다. 대표적으로 엑시 인피니티와 크립토 블레이드, 마이 디파이 펫 등 수집형 게임들이 있다.

2. 사고 경위
1) 보도 내용
2022년 3월 29일 블록체인 기반 게임 엑시 인피니티가 해킹당하는 사건이 발생했다. 해커들이 엑시 인피니티에서 사용되는 암호화폐 이더리움과 USDC를 6억 2,500만 달러어치를 빼돌린 사건이다. 해커들은 엑시 인피니티를 운영하는 스카이마비스가 개발한 사이드체인 로닌 네트워크를 침입해 해킹을 진행했다.
이어서 2022년 4월 14일 미국 재무부는 엑시 인피니티 해킹 사건에 북한 해킹 조직 ‘라자루스’가 연계되어 있다고 발표했다. 그에 대한 조치로 라자루스와 관련되어 있다고 추정되는 암호화폐 이더리움 계좌를 차단하고 중지시키는 조치를 취했다고 밝혔다. 라자루스는 북한 인민군 정찰총국과 연계되었다고 알려진 해커 조직으로써 2014년에 김정은 북한 국무위원장을 조롱한 영화를 제작한 미국 영화사와 소니픽처스를 해킹하는 사건을 일으켰었다.

2) 후속조치
P2E 게임 선두주자 액시 인피티니 해킹 사건이 발생하자 게임 내 취약한 보</td></tr>
</table>

	안체계에 대한 비판이 거세졌다. 해킹을 진행하는 과정에서 해커가 빼돌린 자금을 중앙화 거래소(CEX)에 이체하는 등 미숙한 행동을 보였음에도 불구하고 운영진은 일주일 후에서야 해킹 사실을 알아차렸다. 운영진은 피해를 입은 유저들에게 해킹당한 금액을 반환할 것을 약속했지만 해당 사건을 지켜본 사람들은 취약점 신고보상제도를 뜻하는 버그 바운티(Bug Bounty) 등 해킹을 예방하고 방지할 대책을 내놓길 바라고 있다. **3. 시사점** 암호화폐를 해킹할 경우 토르나도캐시, 스틸엑스 등 믹서와 스왑 프로그램, 탈중앙화 거래소(DEX) 등을 활용하여 자금세탁을 진행하는 경우가 많다. 자금세탁을 진행할 경우 해킹당한 가상화폐를 복구하기 매우 어려우며 범인을 특정하여 추적하는 과정도 많이 힘들어진다. 포트나이츠를 제작한 에픽게임즈의 팀 스위니 CEO는 “NFT가 분산화된 구조를 표방하지만 보안 구조는 취약하다”고 말했으며 필 스펜서 마이크로소프트 게임총괄은 “게임 내 수익 창출 구조도 개발자를 위한 현금 획득 기회일 뿐”이라고 우려를 표명했다. 이번 사건으로 NFT, 메타버스와 맞물려 P2E 시장이 폭발적으로 성장하고 있지만, 이에 비해 보안은 체질개선이 필요하다는 지적에도 힘이 실릴 전망이다. 기존의 시스템에 새로운 기술을 도입하여 편리하고 효율적인 환경을 조성하는 것은 매우 긍정적인 일이지만 사이버 보안은 글로벌 협력이 필요할 만큼 중요한 일이다. 북한과 기타 테러 단체에서 해킹 기술을 이용해 악용하는 사례도 기술이 발전하는 만큼 증가하고 있다. 이를 막기 위해서 정부 기관과 민간 차원에서 보안 관련 전문가를 많이 양성할 필요가 있으며 필요시 ‘사이버 동맹’으로 이어지는 보안 대책이 필요하다. 지금보다 많은 나라가 사이버 위협에 관심을 가져야 할 필요가 있고 일반 이용자들과 기업들도 보안에 대한 예방과 대처법을 고민하고 숙지해야한다.
사고 피해 내역(예상)	**1. 피해 대상:** 베트남 게임 ‘엑시 인피니티’ **2. 피해 내역:** 북한 해킹 조직의 해킹으로 인한 6억 달러 이상의 암호화폐 도난

SW신기술 안전사고 보고서

No. Blockchain-07

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">가상 버킨백 만든 예술가, 에르메스에 피소</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.chosun.com/economy/economy_general/2022/01/22/CXOJEVAQFNE45ALMB5DZDOSUOY/</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2022년 1월 22일</td></tr>
<tr><td>작성자</td><td colspan="3">오경묵</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>프랑스</td><td>사고 유형</td><td>NFT 상표권 및 저작권 논란</td></tr>
<tr><td>사고 일시</td><td>2022년 1월 22일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. NFT
대체 불가능 토큰(Non-Fungible Token, NFT)는 블록체인 기술을 이용해 디지털 형태로 구성된 자산의 소유 증명서 역할을 해주는 토큰이다. 그림과 영상 등 여러 가지의 디지털 자료를 지칭하는 주소를 토큰에 담는 것으로 그 고유성을 지킬 수 있는 원리이다. NFT는 거래내역을 블록체인에 영구적으로 남김으로써 강한 고유성을 가지게 되며 동일품이 존재할 수 없다.
NFT는 2017년도 개념이 정립되어 실험적으로 처음 등장했다. 비트코인이 주목받게 되어 자연스럽게 주목도를 이어받은 블록체인 기술이 발전하면서 NFT도 발전하게 된 것이다. 최초의 NFT는 이미지를 블록체인 안에 기록하고 제작자와 소유권을 증명하는 형태로 사용되었다. 그 후 다양한 프로젝트에서 NFT가 사용되기 시작하며 대중들에게 알려지게 되었다. 주로 하나의 플랫폼 내에서만 그 역할을 할 수 있는 NFT들이 등장했기 때문에 대중들이 많이 접하고 습득할 수 있었지만 큰 의미를 주지는 못했다. 그렇게 NFT가 확산되어 가던 중 디지털 아트에 대한 저작권을 NFT로 구현하기 시작하면서 NFT 붐이 일어났다. NFT는 다양한 형태로 발전하고 있기 때문에 현재 이시간에도 새로운 NFT가 등장하고 있다.

2. 사고 경위
1) 보도 내용
프랑스 명품 브랜드 에르메스에서 ‘가상 버킨백’을 만들어서 판매한 미국 예술가를 상대로 소송을 낸 사건이 발생했다. 2022년 1월 22일 에르메스는 예술가 메이슨 로스차일드에 대해 자사의 상표로 이익을 얻으려 한다는 이유로 미국 뉴욕법원에 고소장을 제출했다. 로스차일드는 지난달 NFT 거래소 오픈시(OpenSea)에 에르메스에서 만들었던 버킨백과 유사한 디지털 작품을 출품했다. 해당 작품은 ‘메타 버킨스(Meta Birkins)’로 실제 에르메스에서 만든 버킨백의 그림파일에 다양한 소재와 색을 입히는 형식으로 작품을 구성했다. 실제 에르메스의 버킨백과 유사한 해당 작품에 많은 사람이 관심을 가졌고</td></tr>
</table>

10억원 이상의 가격이 측정되어 판매될 만큼 인기를 끌었다.
에르메스는 메타 버킨스가 에르메스의 상표권과 저작권을 침해했다고 주장하며 “우리는 수공예정신으로 만든 실물을 중시하기 때문에 NFT 시장에 진출하지 않았고 버킨백의 NFT 제작에 동의한 적이 없다”라고 밝혔다. 이에 로스차일드는 “모피로 뒤덮인 버킨백을 그저 상상하여 예술 작품을 창작했을 뿐이며 위조된 작품이 아니다”라고 대응했다. 하지만 에르메스 측에선 애초 작품의 이름을 메타 버킨스라고 명칭했기에 그 의도가 명확함을 근거로 문제를 제기했다.

2) 로스차일드의 대응
에르메스가 공개적으로 대응에 나서면서 해당 NFT와 작품의 가치가 아예 없는 것이나 마찬가지가 되었다. 구매를 진행한 사람들이 대부분 에르메스의 버킨백에 대한 예술성을 인정하며 구매했기 때문에 그 가치가 훼손된 것이다. 하지만 구매자들과 투자자들은 특별한 조치와 구제 조치를 받지 못하는 상황이다. 피해자들은 당연하게도 해당 작품을 판매한 로스차일드에게 피해보상을 원했지만 이 상황에 대해 로스차일드는 "한번 거래가 완료된 NFT는 이미 완료된 것이기 때문에 메타 버킨스를 구매한 사람들에게 돈을 돌려줄 수 없다”고 밝혔다. 그러면서 "블록체인의 세계는 냉혹하다"고 덧붙였다. 신용카드 등의 결제 수단과 달리 블록체인 상에서 발생한 거래는 불가역적이라는 설명이다.
한 업계 관계자는 “만약 에르메스가 직접 버킨백 NFT를 발행하길 원했다면 메이슨 로스차일드가 에르메스의 권리를 빼앗은 꼴이며 대중들이 메타버킨스를 에르메스가 공식적으로 발행한 것으로 혼동할 만하다”고 전했다. 이어서 "이러한 분쟁은 현재 NFT 시장이 무법지대에 있다는 걸 보여주는 것"이라고 말했다.

3. 시사점
현실 세계와 달리 디지털 세상은 무한에 가까운 복제가 가능하다. 해당 복제품은 원본과 큰 차이와 경계가 존재하지 않고 구분이 매우 힘들다. 이 보장되지 않는 고유성을 지켜주고 증명해주는 것이 NFT라는 기술이다. 사람들은 이 부분에 주목하여 디지털 작품에 대한 많은 관심과 투자를 이어 갔지만 소유에 대한 환상이라는 의견도 나오고 있다. 해당 의견에 대한 근거로 몇 가지가 있는데, 첫 번째로 NFT는 표절이나 해킹 가능성이 있는 불완전한 기술이라는 것이다. 이는 법적으로 보호가 불가능하다는 문제로 이어진다. 두 번째로 해당 사건과 같이 원작자의 허락 없이 등록되는 NFT에 대한 문제가 있다. 그 외에도 NFT에 대한 규제와 인식이 아직 부족한 단계이기 때문에 발생하는 다양한 문제들이 존재하고 있다.
NFT는 고유성을 강조하는 기술이기 때문에 창의적 표현을 더욱 견고하게 사유재산으로 포함하기 위한 방향으로 계속해서 발전해 나갈 것이다. 하지만 저작권에 대한 정확한 규제가 존재하지 않는다면 여러 분야에서 잠깐 나타났다 사라지는 기술이 되어버릴지 모른다. NFT에 대한 인식과 법규의 변화에

	대해 꾸준히 관심을 가지고 다양한 의견을 제시한다면 디지털 기술과 미술의 창작에 있어서 새로운 지평을 열어갈 수 있을 것이다.
사고 피해 내역(예상)	1. **피해 대상**: 명품 브랜드 에르메스 혹은 예술가 메이슨 로스차일드 2. **피해 내역**: 불분명한 NFT 관련 저작권 규제 때문에 생긴 저작권 관련 법정 다툼

SW신기술 안전사고 보고서

No. Blockchain-08

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">크림 파이낸스, 플래시론 공격으로 1500억원 손실</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://www.coindeskkorea.com/news/articleView.html?idxno=75928</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 10월 28일</td></tr>
<tr><td>작성자</td><td colspan="3">김세진</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>미국</td><td>사고 유형</td><td>플래시 론 공격</td></tr>
<tr><td>사고 일시</td><td>2022년 3월 16일</td><td>사고 지속시간</td><td>2일</td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 플래시론 공격
블록체인을 이용한 DeFi는 블록체인의 특성상 신용을 이용한 대출이 불가능하기 때문에 일반적으로 암호화폐를 하나 담보로 삼은 다음 다른 암호화폐를 대출하는 형식을 채용하고 있다. 플래시 론 공격은 이 암호화폐를 담보로 삼는 과정의 허점을 이용해 공격하는 방식이다. 담보를 통해 대출을 진행하기 위해선 해당 담보의 가치를 평가하는 단계를 조작하여 공격을 진행한다. 플래시 론은 사실 공격이 아닌 하나의 대출 방식인데, 일시적으로 담보를 설정하고 대출을 받은 후 대출 받은 암호화폐를 이용한 조작을 통해 이득을 챙긴 후 곧바로 담보를 갚아버리는 형식이다. 플래시 론 공격은 이런 플래시 론을 강제로 가능하게 만드는 공격이다.
순서를 살펴보면, 먼저 암호화폐를 빌릴 때 담보를 설정하는데 이 때 일시적으로 담보의 가치를 조작하여 실제 담보의 가치보다 더 많은 이득을 취하며 대출을 받은 후 대출을 바로 갚아 해당 차익만큼의 이익을 공격자가 가져가는 것이다.

2. 사고 경위
1) 보도 내용
2021년 10월 27일, DeFi 대출 프로젝트 크림 파이낸스(Cream Finanace)에서 플래시 론 공격으로 인해 약 1억 3000만달러의 가상화폐 탈취 피해가 발생했다. 공격자는 주로 크림파이낸스의 유동성공급(LP) 토큰인 크림(CREAM)과 이더리움(ERC-20)기반 토큰들을 탈취했으며, 이 과정에서 68종 토큰을 대상으로 9이더리움(ETH) 이상의 수수료를 사용한 것으로 알려졌다. 해당 사건으로 인해 이더리움 기반으로한 크림 파이낸스의 유동성풀이 동결되었으며 복구는 불가능 하다고 알려졌다. 크림 파이낸스가 외부 공격으로 인해 가상화폐 손해를 보게된 세 번째 사건이다. 크림파이낸스는 2021년 8월 같은 플래시 론 공격으로 1880만 달러를, 2021년 2월 3750만 달러 상당을 피해본 사례가 있다. DeFi 자금손실을 다루는 REKT가 발표한 자료 기준 이번 사건은 역대 디파이 자금탈취 공격 중 세 번째로 큰 규모다.</td></tr>
</table>

	2) 기타 플래시 론 피해 내역 - 2021년 5월 2일, 바이낸스스마트체인 기반 DeFi 프로젝트 스파르탄 프로토콜, 플래시 론 공격으로 인해 3000만달러 피해 - 2021년 5월 20일, 바이낸스스마트체인 기반 이자농사 최적화 서비스 팬케이크버니, 플래시 론 공격으로 인해 코인 가치 99%폭락 - 2022년 5월 28일, 클레이튼 DeFi 프로젝트 팝콘머니, 플래시 론 공격으로 인해 17억원 이상의 담보물 탈취 이 외에도 수많은 피해내역이 있다. 2022년 2분기에만 27건의 플래시 론 공격이 발생하여 약 4천억원의 피해가 발생했으며 1분기 피해액인 1400만 달러의 22배와 같다. 그 외에도 2022년 상반기 블록체인의 허점을 이용한 해킹, 사기, 피싱 등 다양한 공격의 피해액은 2조 6천억원 이상이다. 피해 횟수와 피해액은 점점 증가하고 있으며 그만큼 피해자도 늘어나고 있다. **3. 시사점** 대부분의 업계에서는 이날 발생한 공격의 원인에 대해 플래시 론에 집중하기보다 크림 파이낸스의 스마트 계약 속 취약성에 주목하고 있다. 플래시 론은 사실 기본적으로 자본을 조달하기 위한 수단이었을 뿐 자잘한 프로젝트의 취약성으로 인해 문제가 머니 레고(Money lego)로 이어졌다는 것이다. 머니 레고란, DeFi 시장의 여러 프로토콜이 제공하는 금융 서비스를 마치 레고를 조립하듯 조합하여 수익률을 극대화시키는 작업을 말한다. 플래시 론은 자본이 갖춰지지 못한 사람들에게 대출의 기회를 제공했을 뿐 플래시 론 자체가 취약점이라고 보기엔 어렵기 때문에 플래시 론처럼 여러 DeFi 상품 및 프로토콜에 연결할 수 있는 머니 레고 시스템의 보안 강화가 매우 중요하다고 볼 수 있다. DeFi 산업 차원에서 보안 및 위험원 감사에 대해 표준을 정립해야할 필요가 있다.
사고 피해 내역(예상)	**1. 피해 대상:** DeFi 대출 프로젝트 크림 파이낸스(Cream Finanace) **2. 피해 내역:** 플래시 론 공격으로 인해 약 1억 3000만달러의 가상화폐 탈취 피해 발생

SW신기술 안전사고 보고서

No. Blockchain-09

사고 분야	Blockchain			
사고명	중국, 모든 가상화폐 거래불법화 발표			
사고 보도 정보	사고자료 출처	https://www.yna.co.kr/view/AKR20210924139052083		
	보도 날짜	2021년 9월 24일		
	작성자	조준형		
사고 세부 내용	사고 국가(장소)	중국	사고 유형	가상화폐 규제
	사고 일시	2021년 9월 24일	사고 지속시간	진행 중
사고 내용	**1. 가상화폐** 가상화폐는 지폐나 동전과 같은 실물인 화폐가 아닌 네트워크로 연결된 특정 가상공간에서 활용되는 디지털 화폐 혹은 전자 화폐이다. 그 중 암호화폐는 가상화폐의 일종으로서 비트코인이 암호화폐에 속한다. 현재의 가상화폐는 사람들의 경제 활동에 스며들었으며 시장에서 실제로 활용되기도 하고 있다. 하지만 많은 국가에서 가상화폐에 대한 규제 법안을 내고 있지 못하는 실정이며 어떻게 활용해야 할지조차 정하지 못했다. 가상화폐들은 기능 및 용도가 매우 다양한데, 기술 또한 발전하면서 규제를 정하는 입장에서 갈피를 잡지 못하는 것이 그 원인 중 하나다. 전세계 국가 중 암호화폐 거래를 상대적으로 자유롭게 허용하는 미국, 영국, 프랑스, 독일, 캐나다, 일본 등도 ICO(Initial coin offering, 초기 코인 공개)에 대해서는 강한 규제를 통해 통제하고 있다. 해당 나라들은 암호화폐 거래와 채굴 등으로 인한 소득에 대하여 많은 세금을 부과하고, 이중과세 논란이 있는 부가가치세는 부과하지 않는 식으로 규제를 가하고 있다. 독일은 부가가치세를 부과하다가 유럽사법재판소의 2015년 10월 판결로 철회했다. 암호화폐의 특성상 금융 규제를 피해 가기에 편리하기에 러시아, 이란 등 미국의 금융 제재 대상인 국가들이 국영 암호화폐 발행을 검토하고 있다. 또한 경제가 박살나고 자국 화폐 가치가 쓰레기가 된 국가들이 국영 암호화폐를 발행하기도 한다. **2. 사고 경위** 1) 보도 내용 2021년 9월 24일 중국의 중앙은행인 중국인민은행이 모든 종류의 가상화폐 거래를 불법으로 규정하고 엄격한 단속을 시작하기로 발표했다. 중국인민은행은 “가상화폐는 법정 화폐와 동등한 법적 지위를 보유하지 않는다. 가상화폐 관련 업무 활동은 불법적인 금융 활동에 속한다”라고 밝혔다. 이때 가상화폐 관련 업무에 포함되는 범위는 가상화폐 거래에 정보를 중개하고 정해진 돈을 받는 모든 서비스와 파생상품 거래, 불법 모금도 모두 포함된다. 2020년까지 세계 가상화폐 채굴의 65%를 차지한 중국에서 이러한 파격적인 발표를 하게 되었다.			

중국 국무원이 류허 부총리 주재로 금융안정발전위원회 회의를 연 후 “비트코인 채굴과 거래 행위를 타격하겠다”라고 발언한 배경에는 가상화폐를 이용한 사기, 돈세탁 등과 같은 범죄에 대한 우려와 채굴 과정에서 발생하는 과도한 에너지 사용이 있었다고 예상된다. 중국은 온실가스 배출량 감축에 노력을 기하고 있었지만 채굴에 사용되는 에너지양은 방치할만한 수준이 아니었다고 추측된다. 국무원 발표 이후 네이멍구, 신장 위구르 자치구, 칭하이, 쓰촨, 윈난, 안후이, 허베이성 등 중국 각지에서 비트코인 채굴 단속에 나선 데 이어 이번에 중앙은행 차원에서 전국에 걸친 '가상화폐 거래 전면 불법화' 방침이 발표되기에 이르렀다.

2) 규제의 영향

중국이 가상화폐 채굴을 금지하면서 중국에서 사용되던 채굴기가 세계 각 국으로 퍼져 나갔다. 중국의 가상화폐 채굴 금지 조치 이후 채굴 규모가 큰 14개 암호화폐 채굴 업체들이 중국에서 200만대 이상의 채굴기를 미국과 캐나다, 카자흐스탄, 러시아 등으로 옮겨 갔다.

미국은 재무부, 연준, 증권거래위원회, 상품선물거래위원회 등이 암호화폐 규제를 위한 협의체를 구성했다. 미국은 암호화폐 규제 방안을 담은 보고서를 보고한 후 규제를 발표해 나갈 예정이다. 미국과 중국은 가상화폐을 규제해야 한다는 입장에서는 같은 방향성을 가지고 있다. 그러나 접근 방법과 시장에 대하는 태도는 다르다고 볼 수 있다. 미국은 중국처럼 암호화폐를 직접적으로 탄압하는 방향을 선택하지 않고 암호화폐와 친화적으로 다가가 자신들이 가지고 있는 규제의 틀 안에 집어넣는 것을 목표로 하고 있다. 기존 금융 시스템을 유지하고 돈이 되는 새로운 사업으로써 이용하겠다는 것이다.

하지만 최근 중국 당국의 강력한 가상화폐 채굴, 거래 금지 조치에도 중국의 비트코인 채굴량이 세계적으로 미국에 이은 최상위권에 다시 속하게 되었다고 보도되기도 했다.

3. 시사점

전 세계적으로 유행하는 암호화폐에 대해 각 국의 정부들은 투자 과열을 우려하여 규제 가능성을 열어가고 있으며 실제로 규제하는 나라들과 허용하는 나라들로 구분되어가고 있다.

규제를 찬성하는 측에서는 암호화폐의 실제 기술적 실효성과 통화로서의 가치를 높게 쳐 사용되는 것이 아니라 그저 투기자산으로 활용하여 비이성적으로 활용하고 있다는 의견이다. 암호화폐로 일어난 다단계와 같은 가치 상승이 기술 발전을 보증하지 않았으며 그저 이익을 취하기 위한 도박장으로써 운영되고 있다는 의견도 있다.

규제를 반대하는 측에서는 미래의 기술로 나아가기 전 격동기라고 보며 제도권 내에 잘 흡수되기 위한 규제는 찬성하지만 전면 불법으로 정해버리는 것은 반대한다는 의견을 내세우고 있다. 실제로 이론적으로는 매우 안정성이 보장된 기술이라고 볼 수 있으며 블록체인의 지속적인 보안 강화로 생태계를 잘 유지해 나갈 수 있다는 것이다.

	그 외에도 가상화폐에 대해서 환경문제와 재산권 등 여러 논란이 있다. 중국과 같이 강한 규제만이 정답이 아니다. 정부는 많은 이들에게 혼란을 주지 않으며 피해를 주지 않는 방향으로 규제를 가해야 할 것이며 하나의 사회적 이슈로써 해결해 나가야 할 문제이다.
사고 피해 내역(예상)	1. **피해 대상**: 가상화폐를 이용하는 중국 국민 및 중국의 암호화폐를 이용하던 사람들 2. **피해 내역**: 중국의 강력한 규제로 인해 가상화폐가 금지되고 이용이 불가능해짐.

SW신기술 안전사고 보고서

No. Blockchain-10

<table>
<tr><td>사고 분야</td><td colspan="4">Blockchain</td></tr>
<tr><td>사고명</td><td colspan="4">암호화폐 채굴 장려하던 카자흐스탄, 전력난 심각</td></tr>
<tr><td rowspan="3">사고 보도 정보</td><td>사고자료 출처</td><td colspan="3">https://zdnet.co.kr/view/?no=20211129135259</td></tr>
<tr><td>보도 날짜</td><td colspan="3">2021년 11월 29일</td></tr>
<tr><td>작성자</td><td colspan="3">이정현</td></tr>
<tr><td rowspan="2">사고 세부 내용</td><td>사고 국가(장소)</td><td>카자흐스탄</td><td>사고 유형</td><td>채굴로 인한 전력난</td></tr>
<tr><td>사고 일시</td><td>2022년 11월 28일</td><td>사고 지속시간</td><td></td></tr>
<tr><td>사고 내용</td><td colspan="4">1. 비트코인 채굴과 에너지 소비
비트코인을 채굴할 때 소모되는 에너지가 환경에 매우 안좋은 영향을 미치고 있다는 비판이 커지고 있다. 해당 이슈는 2021년 5월 테슬라 대표 일론 머스크가 트위터에 “암호화폐는 여러 방면에서 좋은 아이디어지만 환경에 매우 큰 피해를 줄 수 있다”라고 발언하며 지원하던 비트코인 결제를 중단하는 사건을 계기로 시작되었다. 해당 발언 이후 비트 코인 가격이 매우 하락하기도 하였으며 많은 사람들이 주목하게 되었기 때문이다.
비트코인이 많은 전기를 사용하는 이유는 작업증명(Proof of work, PoW) 시스템이 기반이 되기 때문이다. 기본적으로 블록체인 네트워크는 10분간의 시간마다 채굴하는 사람 1명이 새로운 블록을 추가할 수 있기 때문에 블록을 추가해 보상을 얻기 위해 24시간 내내 컴퓨터를 켜고 다른 컴퓨터들과 경쟁해야 한다. 이를 위해서 더 많은 장비를 구입해야하고 더 빠르고 에너지 소모가 큰 장비를 갖춰야 하기 때문에 더욱 많은 에너지 소모가 생기는 것이다.

2. 사고 경위
1) 보도 내용
2021년 11월 29일, 정부 차원에서 암호화폐 채굴을 장려하던 카자흐스탄이 심각한 전력난을 겪고 있는 것으로 알려졌다. 카자흐스탄 전력공급회사는 최근 정부에 등록된 채굴업자 50명에게만 전력을 공급하기로 결정했다고 발표했다. 이는 2021년 10월 카자흐스탄 내 석탄 화력 발전소 3곳을 폐쇄된 이후 전력 수급이 원활치 않기 때문이다. 카자흐스탄 에너지부는 2021년 말까지 전력 수요가 작년 대비 1~2% 증가 수준이 아닌 8%까지 급증했다고 추정했다. 카자흐스탄에서는 지난 10월 이후 현재까지 6개 지역에서 정전이 발생했다. 카자흐스탄 정부와 전문가들은 카자흐스탄 지역에서 정부에 등록하지 않고 집이나 공장에서 불법적으로 채굴하는 암호화폐 채굴자들이 급증하면서 정전 사태가 발생하는 것이라고 지적하고 있다. 또, 최근 중국이 암호화폐와의 전쟁을 벌이자 채굴자들이 암호화폐 채굴을 장려하는 카자흐스탄으로 대거 옮긴 까닭도 있다.</td></tr>
</table>

카자흐스탄은 이 같은 전력난을 해결하기 위해 러시아 에너지 회사와 국가 전력망을 보완하는 논의를 시작했으며, 2022년부터는 등록된 채굴자에게 킬로와트시당 1텡게(약 0.0023 달러)의 수수료도 부과할 예정이다. 이에 대해 카자흐스탄 정부가 국가 전력망의 문제를 암호화폐 채굴자들에게 돌리고 있다는 주장도 나오고 있다.

2) 미국과 EU의 반응

비트코인의 전 세계 연간 전력 소비량은 약 130테라와트로 추정되는데, 노르웨이나 스웨덴 같은 국가의 1년 총 전력 소비량과 맞먹는다. 여기에 두 번째로 규모가 큰 이더리움의 전력 소비량을 더하면, 무려 300테라와트에 달한다. 영국의 1년 전력 소비량과 비교할 만한 수치이다. 비트코인과 이더리움을 채굴하는 데 드는 전력은 탄소 배출에도 심각한 영향을 미치는데, 2021년에만 8,000만 톤의 Co2를 배출한 것으로 추정된다. 암호화폐의 전력 소비량과 전 세계적인 지속 가능성 목표를 비교해보면, 많은 국가에서 암호화폐 채굴에 대한 무비판적인 수용을 중단한 것은 당연한 대응이다. 특히 미국의 대응이 두드러진데, 미국은 중국의 암호화폐 채굴 금지 이후 암호화폐 채굴의 중심지로 성장했다. 케임브리지 대학의 채굴 지도에 따르면, 비트코인 채굴에 사용되는 전 세계 컴퓨팅 성능 중 미국이 차지하는 비중은 2019년 1월 4%에서 2022년 1월 38%로 증가했다.

이에 미국의 첫 번째 대응으로 민주당은 EPA와 에너지부에 공개 서한을 보내 암호화폐 채굴 업체의 탄소 배출량과 전력 소비량을 공개할 것으로 요청했다. 유럽연합에서도 암호화폐 채굴과 거래는 더 이상 쉽게 받아들여지지 않는다. 6월 말에 EU 의회와 이사회 의장국은 MiCA(Markets in Crypto-Assets)란 새로운 법안에 잠정 합의했다.

3. 시사점

비트코인이 환경파괴의 주범으로 꼽히는 이유는 전기의 출처다. 비트코인 산업은 방대한 전력을 소모하는데, 사용되는 전력의 대부분은 석탄화력발전소와 같은 재생불가능한 곳에서 생성된다는 게 세간의 인식이다. 비트코인 산업에서 사용하는 화석 연료가 엄청난 양의 이산화탄소를 대기 중으로 방출하고, 이는 기후변화에 악영향을 미친다는 설명이다. 특히 2017년 가격 상승 이후 비트코인 네트워크가 약 4배 증가하면서 비트코인에 대한 높은 에너지 수요는 더욱더 우려를 샀다. 비트코인 채굴 자체가 무조건적으로 탄소 배출 증가로 이어지지는 않지만, 비트코인이 소비하는 에너지 중 39%만이 탄소중립적인 에너지라는 점이 우려의 원인이다.

하지만 업계에서는 비트코인이 사용하는 전력의 출처 중 친환경적인 요소가 존재하며, 향후 더 친환경적으로 생산할 수 있는 가능성이 크다는 점을 봐야 한다고 강조하고 있다. 비트코인의 채굴자들이 쓰는 전기의 양에 주목하거나 이 전기가 친환경적이지 않다고 단정하기 전에, 이 전기를 어디에서 공급하고, 탄소배출량에 어떤 영향을 주는지에 대한 논의가 중요하다는 설명이다. 업계에서는 비트코인과 암호화폐 채굴을 보다 지속가능하고 환경 친화적으로

	만들기 위한 시도가 계속되고 있다. 미래 핵심기술 중 하나인 블록체인이 완전히 정착하기 위해서는 환경오염을 촉발하지 않는 형태로 블록체인이 운영될 수 있도록 기술적인 측면에서도 꾸준한 개발이 필요할 것이다.
사고 **피해 내역(예상)**	1. **피해 대상**: 카자흐스탄 2. **피해 내역**: 비트코인 채굴을 위한 에너지 소모로 인한 정전과 전력난

참 고 문 헌

[국내 문헌]

KISDI (2021), 「메타버스 시대의 디지털 플랫폼 규제」

[해외 문헌]

NTSB (2019), 「Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian」

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

R. Chillarege, I.S. Bhandari, J.K. Chaar, M.J. Halliday, D.S. Moebus, B.K. Ray, and M.-Y. Wong. Orthogonal defect classification-a concept for in-process measurements. IEEE Transactions on Software Engineering, 18(11):943-956, 1992.

Richard A. DeMillo and Aditya P. Mathur. A grammar based fault classification scheme and its application to the classification of the errors of TEX. Technical report, Software Engineering Research Center and Department of Computer Sciences, Purdue University, 1995.

Marc Eisenstadt. My hairiest bug war stories. Commun. ACM, 40(4):30-37, 1997.

European Commission, Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Markets in Crypto-assets, and amending Directive (EU) 2019/1937

European Parliament, Blockchain and the General Data Protection Regulation, European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 634.445, July 2019

Jim Gray. Why do computers stop and what can be done about it In Proceedings of Symposium on Reliability in Distributed Software and Database Systems (SRDS-5),

pages 3-12. IEEE CS Press, 1986.

Huang B, Zhang H, Lu M. Software FMEA approach based on failure modes database. Proceedings of the eighth international conference on reliability. Maintainability and Safety; 2009. p. 749-53.

Integrated Innovation Strategy Promotion Council, Japan, AI Stratedgy 2022: AI for all people, industries, regions, and governments. 2019

Donald E. Knuth. The errors of TEX. Software Practice and Experience, 19(7):607-685, July 1989.

Nancy Engineering a Safer World: Systems Thinking Applied to Safety (Engineering Systems), 2016

Li B. Integrating software into PRA (Probabilistic risk assessment) [Monograph]. College Park, MD: University of Maryland; 2004.

Li B, Li M, Ghose S, Smidts C. Integrating software into PRA. Proceedings of the Fourteenth International Symposium on Software Reliability Engineering, ISSRE 2003. 2003. p. 457-67.

Li B, Li M, Smidts C. Integrating software into PRA: a test-based approach. Risk Anal 2005;25:1061-77.

OECD. Failure modes taxonomy for reliability assessment of digital I&C systems for PRA. Paris, France: Organisation for Economic Co-operation and Development, Nuclear Energy Agency; 2014. p. 1-135.

Park GY, Kim DH, Lee DY. Software FMEA analysis for safety-related application software. Ann Nucl Energy 2014;70:96-102.

Jan Ploski, Matthias Rohr, Peter Schwenkenberg, and Wilhelm Hasselbring, "Research Issues in Software Fault Categorization," ACM SIGSOFT Software Engineering Notes, Vol. 32, No. 6, pp.1-8, November 2007.

Prasanna KN, Gokhale SA, Agarwal R, Chetwani RR, Ravindra M, Bharadwaj KM. Application of software failure mode and effect analysis for on-board software.

Proceedings of the international conference on advances in computing, communications and informatics (ICACCI). 2014. p. 684-8.

Ristord L, Esmenjaud C. FMEA performed on the SPINLINE3 operational system software as part of the TIHANGE 1 NIS refurbishment safety case. Proceedings of the CNRA/CSNI workshop on licensing and operating experience of computer-based I&C systems. 2002. p. 37-50. Czech Republic: NEA/CSN/OECD.

Stadler JJ, Seidl NJ. Software failure modes and effects analysis. Proceedings of the fifty-ninth annual reliability and maintainability symposium, RAMS 2013. Institute of Electrical and Electronics Engineers Inc.; 2013. p. 1-5.

Thieme, Christoph A., et al. "Incorporating software failure in risk analysis-Part 1: Software functional failure mode classification." Reliability Engineering & System Safety 197 (2020): 106803.

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.

비매품/무료
95060
9 791191 525274
ISBN 979-11-91525-27-4 (PDF)

[소프트웨어정책연구소]에 의해 작성된 **[SPRI 보고서]**는 공공저작물 자유이용허락 표시기준 제 4유형**(출처표시-상업적이용금지-변경금지)**에 따라 이용할 수 있습니다.
(출처를 밝히면 자유로운 이용이 가능하지만, 영리목적으로 이용할 수 없고, 변경 없이 그대로 이용해야 합니다.)